

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
25 October 2001 (25.10.2001)

PCT

(10) International Publication Number  
**WO 01/79472 A2**

- (51) International Patent Classification<sup>7</sup>: C12N 9/00 (74) Agent: MARSH, David, R.; Arnold & Porter, 555 12th Street, NW, Washington, DC 20004-1206 (US).
- (21) International Application Number: PCT/US01/12334
- (22) International Filing Date: 13 April 2001 (13.04.2001) (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
09/549,848 14 April 2000 (14.04.2000) US  
09/688,069 14 October 2000 (14.10.2000) US
- (71) Applicant: MONSANTO TECHNOLOGY LLP [US/US]; 800 North Lindbergh Blvd., St. Louis, MO 63167 (US).
- (72) Inventors: SUBRAMANIAM, Sai, S.; 33 Dinsmore Avenue #304, Framingham, MA 01702 (US). SLATER, Steven, C.; 21 Brucewood Road, Acton, MA 01720 (US). KARBERG, Katherine; 37 Regent Street #3, Cambridge, MA 02140 (US). CHEN, Ridong; 11131-I Westport Station Drive, Maryland Heights, MO 63043 (US). VALENTIN, Henry, E.; 37 Regent Street #3, Cambridge, MA 02140 (US). WONG, Yun-Hua, Huang; 14043 Forest Crest Drive, Chesterfield, MO 63017 (US).
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:**  
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*



WO 01/79472 A2

(54) Title: NUCLEIC ACID SEQUENCES TO PROTEINS INVOLVED IN TOCOPHEROL SYNTHESIS

(57) Abstract: Nucleic acid sequences and methods are provided for producing plants and seeds having altered tocopherol content and compositions. The methods find particular use in increasing the tocopherol levels in plants, and in providing desirable tocopherol compositions in a host plant cell.

5                   **NUCLEIC ACID SEQUENCES TO PROTEINS INVOLVED IN  
                          TOCOPHEROL SYNTHESIS**

**INTRODUCTION**

**TECHNICAL FIELD**

10           The present invention is directed to nucleic acid and amino acid sequences and constructs, and methods related thereto.

**BACKGROUND**

          Isoprenoids are ubiquitous compounds found in all living organisms. Plants synthesize a diverse array of greater than 22,000 isoprenoids (Connolly and Hill  
15   (1992) *Dictionary of Terpenoids*, Chapman and Hall, New York, NY). In plants, isoprenoids play essential roles in particular cell functions such as production of sterols, contributing to eukaryotic membrane architecture, acyclic polyprenoids found in the side chain of ubiquinone and plastoquinone, growth regulators like abscisic acid, gibberellins, brassinosteroids or the photosynthetic pigments chlorophylls and  
20   carotenoids. Although the physiological role of other plant isoprenoids is less evident, like that of the vast array of secondary metabolites, some are known to play key roles mediating the adaptative responses to different environmental challenges. In spite of the remarkable diversity of structure and function, all isoprenoids originate from a single metabolic precursor, isopentenyl diphosphate (IPP) (Wright, (1961) *Annu. Rev. Biochem.* 20:525-548; and Spurgeon and Porter, (1981) in Biosynthesis of Isoprenoid Compounds., Porter and Spurgeon eds (John Wiley, New York) Vol. 1, pp1-46).  
25

          A number of unique and interconnected biochemical pathways derived from the isoprenoid pathway leading to secondary metabolites, including tocopherols, exist in chloroplasts of higher plants. Tocopherols not only perform vital functions in

plants, but are also important from mammalian nutritional perspectives. In plastids, tocopherols account for up to 40% of the total quinone pool.

Tocopherols and tocotrienols (unsaturated tocopherol derivatives) are well known antioxidants, and play an important role in protecting cells from free radical damage, and in the prevention of many diseases, including cardiac disease, cancer, cataracts, retinopathy, Alzheimer's disease, and neurodegeneration, and have been shown to have beneficial effects on symptoms of arthritis, and in anti-aging. Vitamin E is used in chicken feed for improving the shelf life, appearance, flavor, and oxidative stability of meat, and to transfer tocopherols from feed to eggs. Vitamin E has been shown to be essential for normal reproduction, improves overall performance, and enhances immunocompetence in livestock animals. Vitamin E supplement in animal feed also imparts oxidative stability to milk products.

The demand for natural tocopherols as supplements has been steadily growing at a rate of 10-20% for the past three years. At present, the demand exceeds the supply for natural tocopherols, which are known to be more biopotent than racemic mixtures of synthetically produced tocopherols. Naturally occurring tocopherols are all *d*-stereoisomers, whereas synthetic  $\alpha$ -tocopherol is a mixture of eight *d,l*- $\alpha$ -tocopherol isomers, only one of which (12.5%) is identical to the natural *d*- $\alpha$ -tocopherol. Natural *d*- $\alpha$ -tocopherol has the highest vitamin E activity (1.49 IU/mg) when compared to other natural tocopherols or tocotrienols. The synthetic  $\alpha$ -tocopherol has a vitamin E activity of 1.1 IU/mg. In 1995, the worldwide market for raw refined tocopherols was \$1020 million; synthetic materials comprised 85-88% of the market, the remaining 12-15% being natural materials. The best sources of natural tocopherols and tocotrienols are vegetable oils and grain products. Currently, most of the natural Vitamin E is produced from  $\gamma$ -tocopherol derived from soy oil processing, which is subsequently converted to  $\alpha$ -tocopherol by chemical modification ( $\alpha$ -tocopherol exhibits the greatest biological activity).

Methods of enhancing the levels of tocopherols and tocotrienols in plants, especially levels of the more desirable compounds that can be used directly, without

chemical modification, would be useful to the art as such molecules exhibit better functionality and bioavailability.

In addition, methods for the increased production of other isoprenoid derived compounds in a host plant cell is desirable. Furthermore, methods for the production of particular isoprenoid compounds in a host plant cell is also needed.

#### SUMMARY OF THE INVENTION

The present invention is directed to sequences to proteins involved in tocopherol synthesis. The polynucleotides and polypeptides of the present invention include those derived from prokaryotic and eukaryotic sources.

Thus, one aspect of the present invention relates to prenyltransferase, and in particular to isolated polynucleotide sequences encoding prenyltransferase proteins and polypeptides related thereto. In particular, isolated nucleic acid sequences encoding prenyltransferase proteins from bacterial and plant sources are provided.

In another aspect, the present invention provides isolated polynucleotide sequences encoding tocopherol cyclase, and polypeptides related thereto. In particular, isolated nucleic acid sequences encoding tocopherol cyclase proteins from bacterial and plant sources are provided.

Another aspect of the present invention relates to oligonucleotides which include partial or complete prenyltransferase or tocopherol cyclase encoding sequences.

It is also an aspect of the present invention to provide recombinant DNA constructs which can be used for transcription or transcription and translation (expression) of prenyltransferase or tocopherol cyclase. In particular, constructs are provided which are capable of transcription or transcription and translation in host cells.

In another aspect of the present invention, methods are provided for production of prenyltransferase or tocopherol cyclase in a host cell or progeny thereof. In particular, host cells are transformed or transfected with a DNA construct which can be used for transcription or transcription and translation of



prenyltransferase or tocopherol cyclase. The recombinant cells which contain prenyltransferase or tocopherol cyclase are also part of the present invention.

In a further aspect, the present invention relates to methods of using polynucleotide and polypeptide sequences to modify the tocopherol content of host  
5 cells, particularly in host plant cells. Plant cells having such a modified tocopherol content are also contemplated herein. Methods and cells in which both prenyltransferase and tocopherol cyclase are expressed in a host cell are also part of the present invention.

The modified plants, seeds and oils obtained by the expression of the  
10 prenyltransferase or tocopherol cyclase are also considered part of the invention.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

Figure 1 provides an amino acid sequence alignment between ATPT2, ATPT3, ATPT4, ATPT8, and ATPT12 are performed using ClustalW.

Figure 2 provides a schematic picture of the expression construct pCGN10800.

15 Figure 3 provides a schematic picture of the expression construct pCGN10801.

Figure 4 provides a schematic picture of the expression construct pCGN10803.

Figure 5 provides a schematic picture of the construct pCGN10806.

Figure 6 provides a schematic picture of the construct pCGN10807.

Figure 7 provides a schematic picture of the construct pCGN10808.

20 Figure 8 provides a schematic picture of the expression construct pCGN10809.

Figure 9 provides a schematic picture of the expression construct pCGN10810.

Figure 10 provides a schematic picture of the expression construct pCGN10811.

Figure 11 provides a schematic picture of the expression construct pCGN10812.

Figure 12 provides a schematic picture of the expression construct pCGN10813.

25 Figure 13 provides a schematic picture of the expression construct pCGN10814.

Figure 14 provides a schematic picture of the expression construct pCGN10815.

Figure 15 provides a schematic picture of the expression construct pCGN10816.

Figure 16 provides a schematic picture of the expression construct pCGN10817.

Figure 17 provides a schematic picture of the expression construct pCGN10819.

Figure 18 provides a schematic picture of the expression construct pCGN10824.

Figure 19 provides a schematic picture of the expression construct pCGN10825.

Figure 20 provides a schematic picture of the expression construct pCGN10826.

Figure 21 provides an amino acid sequence alignment using ClustalW between the  
5 *Synechocystis* prenyltransferase sequences.

Figure 22 provides an amino acid sequence of the ATPT2, ATPT3, ATPT4, ATPT8, and ATPT12 protein sequences from *Arabidopsis* and the slr1736, slr0926, slr11899, slr0056, and the slr1518 amino acid sequences from *Synechocystis*.

Figure 23 provides the results of the enzymatic assay from preparations of  
10 wild type *Synechocystis* strain 6803, and *Synechocystis* slr1736 knockout.

Figure 24 provides bar graphs of HPLC data obtained from seed extracts of transgenic *Arabidopsis* containing pCGN10822, which provides of the expression of the ATPT2 sequence, in the sense orientation, from the napin promoter. Provided are graphs for alpha, gamma, and delta tocopherols, as well as total tocopherol for 22  
15 transformed lines, as well as a nontransformed (wildtype) control.

Figure 25 provides a bar graph of HPLC analysis of seed extracts from *Arabidopsis* plants transformed with pCGN10803 (35S-ATPT2, in the antisense orientation), pCGN10822 (line 1625, napin ATPT2 in the sense orientation), pCGN10809 (line 1627, 35S-ATPT3 in the sense orientation), a nontransformed (wt)  
20 control, and an empty vector transformed control.

Figure 26 shows total tocopherol levels measured in T# *Arabidopsis* seed of line.

Figure 27 shows total tocopherol levels measured in T# *Arabidopsis* seed of line.

Figure 28 shows total tocopherol levels measured in developing canola seed of  
25 line 10822-1.

Figure 29: shows results of phytyl prenyltransferase activity assay using *Synechocystis* wild type and slr1737 knockout mutant membrane preparations.

Figure 30 is the chromatograph from an HPLC analysis of *Synechocystis* extracts.

Figure 31 is a sequence alignment of the *Arabidopsis* homologue with the sequence of the public database.

5 Figure 32 shows the results of hydropathic analysis of slr1737

Figure 33 shows the results of hydropathic analysis of the *Arabidopsis* homologue of slr1737.

Figure 34 shows the catalytic mechanism of various cyclase enzymes

10 Figure 35 is a sequence alignment of slr1737, slr1737 *Arabidopsis* homologue and the *Arabidopsis* chalcone isomerase.

#### DETAILED DESCRIPTION OF THE INVENTION

The present invention provides, *inter alia*, compositions and methods for altering (for example, increasing and decreasing) the tocopherol levels and/or modulating their ratios in host cells. In particular, the present invention provides  
15 polynucleotides, polypeptides, and methods of use thereof for the modulation of tocopherol content in host plant cells.

The biosynthesis of  $\alpha$ -tocopherol in higher plants involves condensation of homogentisic acid and phytylpyrophosphate to form 2-methyl-6 phytylbenzoquinol that can, by cyclization and subsequent methylations (Fiedler et al., 1982, *Planta*, 155:  
20 511-515, Soll et al., 1980, *Arch. Biochem. Biophys.* 204: 544-550, Marshall et al., 1985 *Phytochem.*, 24: 1705-1711, all of which are herein incorporated by reference in their entirety), form various tocopherols.

The *Arabidopsis pds2* mutant identified and characterized by Norris *et al.* (1995), is deficient in tocopherol and plastoquinone-9 accumulation. Further genetic  
25 and biochemical analysis suggested that the protein encoded by *PDS2* may be responsible for the prenylation of homogentisic acid. The *PDS2* locus identified by Norris *et al.* (1995) has been hypothesized to possibly encode the tocopherol phytyl-prenyltransferase, as the *pds2* mutant fails to accumulate tocopherols.

Norris *et al.* (1995) determined that in *Arabidopsis pds2* lies at the top of chromosome 3, approximately 7 centimorgans above long hypocotyl2, based on the genetic map. ATPT2 is located on chromosome 2 between 36 and 41 centimorgans, lying on BAC F19F24, indicating that ATPT2 does not correspond to *PDS2*. Thus, it is an aspect of the present invention to provide novel polynucleotides and polypeptides involved in the prenylation of homogentisic acid. This reaction may be a rate limiting step in tocopherol biosynthesis, and this gene has yet to be isolated.

U.S. Patent No. 5,432,069 describes the partial purification and characterization of tocopherol cyclase from *Chlorella protothecoides*, *Dunaliella salina* and wheat. The cyclase described as being glycine rich, water soluble and with a predicted MW of 48-50kDa. However, only limited peptide fragment sequences were available.

In one aspect, the present invention provides polynucleotide and polypeptide sequences involved in the prenylation of straight chain and aromatic compounds. Straight chain prenyltransferases as used herein comprises sequences which encode proteins involved in the prenylation of straight chain compounds, including, but not limited to, geranyl geranyl pyrophosphate and farnesyl pyrophosphate. Aromatic prenyltransferases, as used herein, comprises sequences which encode proteins involved in the prenylation of aromatic compounds, including, but not limited to, menaquinone, ubiquinone, chlorophyll, and homogentisic acid. The prenyltransferase of the present invention preferably prenylates homogentisic acid.

In another aspect, the invention provides polynucleotide and polypeptide sequences to tocopherol cyclization enzymes. The 2,3-dimethyl-5-phytylplastoquinol cyclase (tocopherol cyclase) is responsible for the cyclization of 2,3-dimethyl-5-phytylplastoquinol to tocopherol.

#### **Isolated Polynucleotides, Proteins, and Polypeptides**

A first aspect of the present invention relates to isolated prenyltransferase polynucleotides. Another aspect of the present invention relates to isolated tocopherol cyclase polynucleotides. The polynucleotide sequences of the present invention

include isolated polynucleotides that encode the polypeptides of the invention having a deduced amino acid sequence selected from the group of sequences set forth in the Sequence Listing and to other polynucleotide sequences closely related to such sequences and variants thereof.

5           The invention provides a polynucleotide sequence identical over its entire length to each coding sequence as set forth in the Sequence Listing. The invention also provides the coding sequence for the mature polypeptide or a fragment thereof, as well as the coding sequence for the mature polypeptide or a fragment thereof in a reading frame with other coding sequences, such as those encoding a leader or  
10   secretory sequence, a pre-, pro-, or prepro- protein sequence. The polynucleotide can also include non-coding sequences, including for example, but not limited to, non-coding 5' and 3' sequences, such as the transcribed, untranslated sequences, termination signals, ribosome binding sites, sequences that stabilize mRNA, introns, polyadenylation signals, and additional coding sequence that encodes additional  
15   amino acids. For example, a marker sequence can be included to facilitate the purification of the fused polypeptide. Polynucleotides of the present invention also include polynucleotides comprising a structural gene and the naturally associated sequences that control gene expression.

The invention also includes polynucleotides of the formula:



wherein, at the 5' end, X is hydrogen, and at the 3' end, Y is hydrogen or a metal,  $R_1$  and  $R_3$  are any nucleic acid residue,  $n$  is an integer between 1 and 3000, preferably between 1 and 1000 and  $R_2$  is a nucleic acid sequence of the invention, particularly a nucleic acid sequence selected from the group set forth in the Sequence Listing and  
25   preferably those of SEQ ID NOs: 1, 3, 5, 7, 8, 10, 11, 13-16, 18, 23, 29, 36, and 38. In the formula,  $R_2$  is oriented so that its 5' end residue is at the left, bound to  $R_1$ , and its 3' end residue is at the right, bound to  $R_3$ . Any stretch of nucleic acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer.

The invention also relates to variants of the polynucleotides described herein that encode for variants of the polypeptides of the invention. Variants that are fragments of the polynucleotides of the invention can be used to synthesize full-length polynucleotides of the invention. Preferred embodiments are polynucleotides  
5 encoding polypeptide variants wherein 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues of a polypeptide sequence of the invention are substituted, added or deleted, in any combination. Particularly preferred are substitutions, additions, and deletions that are silent such that they do not alter the properties or activities of the polynucleotide or polypeptide.

10 Further preferred embodiments of the invention that are at least 50%, 60%, or 70% identical over their entire length to a polynucleotide encoding a polypeptide of the invention, and polynucleotides that are complementary to such polynucleotides. More preferable are polynucleotides that comprise a region that is at least 80% identical over its entire length to a polynucleotide encoding a polypeptide of the  
15 invention and polynucleotides that are complementary thereto. In this regard, polynucleotides at least 90% identical over their entire length are particularly preferred, those at least 95% identical are especially preferred. Further, those with at least 97% identity are highly preferred and those with at least 98% and 99% identity are particularly highly preferred, with those at least 99% being the most highly  
20 preferred.

Preferred embodiments are polynucleotides that encode polypeptides that retain substantially the same biological function or activity as the mature polypeptides encoded by the polynucleotides set forth in the Sequence Listing.

The invention further relates to polynucleotides that hybridize to the above-  
25 described sequences. In particular, the invention relates to polynucleotides that hybridize under stringent conditions to the above-described polynucleotides. As used herein, the terms "stringent conditions" and "stringent hybridization conditions" mean that hybridization will generally occur if there is at least 95% and preferably at least 97% identity between the sequences. An example of stringent hybridization

conditions is overnight incubation at 42°C in a solution comprising 50% formamide, 5x SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 micrograms/milliliter denatured, sheared salmon sperm DNA, followed by washing the hybridization support in 0.1x SSC at approximately 65°C. Other hybridization and wash conditions are well known and are exemplified in Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor, NY (1989), particularly Chapter 11.

The invention also provides a polynucleotide consisting essentially of a polynucleotide sequence obtainable by screening an appropriate library containing the complete gene for a polynucleotide sequence set forth in the Sequence Listing under stringent hybridization conditions with a probe having the sequence of said polynucleotide sequence or a fragment thereof; and isolating said polynucleotide sequence. Fragments useful for obtaining such a polynucleotide include, for example, probes and primers as described herein.

As discussed herein regarding polynucleotide assays of the invention, for example, polynucleotides of the invention can be used as a hybridization probe for RNA, cDNA, or genomic DNA to isolate full length cDNAs or genomic clones encoding a polypeptide and to isolate cDNA or genomic clones of other genes that have a high sequence similarity to a polynucleotide set forth in the Sequence Listing. Such probes will generally comprise at least 15 bases. Preferably such probes will have at least 30 bases and can have at least 50 bases. Particularly preferred probes will have between 30 bases and 50 bases, inclusive.

The coding region of each gene that comprises or is comprised by a polynucleotide sequence set forth in the Sequence Listing may be isolated by screening using a DNA sequence provided in the Sequence Listing to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the invention is then used to screen a library of cDNA, genomic DNA or mRNA to identify members of the library which hybridize to the probe. For



example, synthetic oligonucleotides are prepared which correspond to the prenyltransferase or tocopherol cyclase EST sequences. The oligonucleotides are used as primers in polymerase chain reaction (PCR) techniques to obtain 5' and 3' terminal sequence of prenyltransferase or tocopherol cyclase genes. Alternatively, where oligonucleotides of low degeneracy can be prepared from particular prenyltransferase or tocopherol cyclase peptides, such probes may be used directly to screen gene libraries for prenyltransferase or tocopherol cyclase gene sequences. In particular, screening of cDNA libraries in phage vectors is useful in such methods due to lower levels of background hybridization.

Typically, a prenyltransferase or tocopherol cyclase sequence obtainable from the use of nucleic acid probes will show 60-70% sequence identity between the target prenyltransferase or tocopherol cyclase sequence and the encoding sequence used as a probe. However, lengthy sequences with as little as 50-60% sequence identity may also be obtained. The nucleic acid probes may be a lengthy fragment of the nucleic acid sequence, or may also be a shorter, oligonucleotide probe. When longer nucleic acid fragments are employed as probes (greater than about 100 bp), one may screen at lower stringencies in order to obtain sequences from the target sample which have 20-50% deviation (i.e., 50-80% sequence homology) from the sequences used as probe. Oligonucleotide probes can be considerably shorter than the entire nucleic acid sequence encoding an prenyltransferase or tocopherol cyclase enzyme, but should be at least about 10, preferably at least about 15, and more preferably at least about 20 nucleotides. A higher degree of sequence identity is desired when shorter regions are used as opposed to longer regions. It may thus be desirable to identify regions of highly conserved amino acid sequence to design oligonucleotide probes for detecting and recovering other related prenyltransferase or tocopherol cyclase genes. Shorter probes are often particularly useful for polymerase chain reactions (PCR), especially when highly conserved sequences can be identified. (*See, Gould, et al., PNAS USA (1989) 86:1934-1938.*)

Another aspect of the present invention relates to prenyltransferase or tocopherol cyclase polypeptides. Such polypeptides include isolated polypeptides set forth in the Sequence Listing, as well as polypeptides and fragments thereof, particularly those polypeptides which exhibit prenyltransferase or tocopherol cyclase activity and also those polypeptides which have at least 50%, 60% or 70% identity, preferably at least 80% identity, more preferably at least 90% identity, and most preferably at least 95% identity to a polypeptide sequence selected from the group of sequences set forth in the Sequence Listing, and also include portions of such polypeptides, wherein such portion of the polypeptide preferably includes at least 30 amino acids and more preferably includes at least 50 amino acids.

"Identity", as is well understood in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as determined by the match between strings of such sequences. "Identity" can be readily calculated by known methods including, but not limited to, those described in *Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York (1988); *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I*, Griffin, A.M. and Griffin, H.G., eds., Humana Press, New Jersey (1994); *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press (1987); *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., Stockton Press, New York (1991); and Carillo, H., and Lipman, D., *SIAM J Applied Math*, 48:1073 (1988). Methods to determine identity are designed to give the largest match between the sequences tested. Moreover, methods to determine identity are codified in publicly available programs. Computer programs which can be used to determine identity between two sequences include, but are not limited to, GCG (Devereux, J., et al., *Nucleic Acids Research* 12(1):387 (1984); suite of five BLAST programs, three designed for nucleotide sequences queries (BLASTN, BLASTX, and TBLASTX) and two

designed for protein sequence queries (BLASTP and TBLASTN) (Coulson, *Trends in Biotechnology*, 12: 76-80 (1994); Birren, *et al.*, *Genome Analysis*, 1: 543-559 (1997)). The BLAST X program is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, S., *et al.*, NCBI NLM NIH, Bethesda, MD 20894; Altschul, S., *et al.*, *J. Mol. Biol.*, 215:403-410 (1990)). The well known Smith Waterman algorithm can also be used to determine identity.

Parameters for polypeptide sequence comparison typically include the following:

Algorithm: Needleman and Wunsch, *J. Mol. Biol.* 48:443-453 (1970)  
 10 Comparison matrix: BLOSSUM62 from Hentikoff and Hentikoff, *Proc. Natl. Acad. Sci USA* 89:10915-10919 (1992)

Gap Penalty: 12

Gap Length Penalty: 4

A program which can be used with these parameters is publicly available as  
 15 the "gap" program from Genetics Computer Group, Madison Wisconsin. The above parameters along with no penalty for end gap are the default parameters for peptide comparisons.

Parameters for polynucleotide sequence comparison include the following:

Algorithm: Needleman and Wunsch, *J. Mol. Biol.* 48:443-453 (1970)

20 Comparison matrix: matches = +10; mismatches = 0

Gap Penalty: 50

Gap Length Penalty: 3

A program which can be used with these parameters is publicly available as  
 the "gap" program from Genetics Computer Group, Madison Wisconsin. The above  
 25 parameters are the default parameters for nucleic acid comparisons.

The invention also includes polypeptides of the formula:



wherein, at the amino terminus, X is hydrogen, and at the carboxyl terminus, Y is hydrogen or a metal,  $R_1$  and  $R_3$  are any amino acid residue, n is an integer between 1

and 1000, and  $R_2$  is an amino acid sequence of the invention, particularly an amino acid sequence selected from the group set forth in the Sequence Listing and preferably those encoded by the sequences provided in SEQ ID NOs: 2, 4, 6, 9, 12, 17, 19-22, 24-28, 30, 32-35, 37, and 39. In the formula,  $R_2$  is oriented so that its amino terminal  
5 residue is at the left, bound to  $R_1$ , and its carboxy terminal residue is at the right, bound to  $R_3$ . Any stretch of amino acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer.

Polypeptides of the present invention include isolated polypeptides encoded by  
10 a polynucleotide comprising a sequence selected from the group of a sequence contained in the Sequence Listing set forth herein .

The polypeptides of the present invention can be mature protein or can be part of a fusion protein.

Fragments and variants of the polypeptides are also considered to be a part of  
15 the invention. A fragment is a variant polypeptide which has an amino acid sequence that is entirely the same as part but not all of the amino acid sequence of the previously described polypeptides. The fragments can be "free-standing" or comprised within a larger polypeptide of which the fragment forms a part or a region, most preferably as a single continuous region. Preferred fragments are biologically  
20 active fragments which are those fragments that mediate activities of the polypeptides of the invention, including those with similar activity or improved activity or with a decreased activity. Also included are those fragments that antigenic or immunogenic in an animal, particularly a human.

Variants of the polypeptide also include polypeptides that vary from the  
25 sequences set forth in the Sequence Listing by conservative amino acid substitutions, substitution of a residue by another with like characteristics. In general, such substitutions are among Ala, Val, Leu and Ile; between Ser and Thr; between Asp and Glu; between Asn and Gln; between Lys and Arg; or between Phe and Tyr.

Particularly preferred are variants in which 5 to 10; 1 to 5; 1 to 3 or one amino acid(s) are substituted, deleted, or added, in any combination.

5 Variants that are fragments of the polypeptides of the invention can be used to produce the corresponding full length polypeptide by peptide synthesis. Therefore, these variants can be used as intermediates for producing the full-length polypeptides of the invention.

The polynucleotides and polypeptides of the invention can be used, for example, in the transformation of host cells, such as plant host cells, as further discussed herein.

10 The invention also provides polynucleotides that encode a polypeptide that is a mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids within the mature polypeptide (for example, when the mature form of the protein has more than one polypeptide chain). Such sequences can, for example, play a role in the processing of a protein from a precursor to a mature form, allow protein  
15 transport, shorten or lengthen protein half-life, or facilitate manipulation of the protein in assays or production. It is contemplated that cellular enzymes can be used to remove any additional amino acids from the mature protein.

A precursor protein, having the mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. The inactive  
20 precursors generally are activated when the prosequences are removed. Some or all of the prosequences may be removed prior to activation. Such precursor protein are generally called proproteins.

#### **Plant Constructs and Methods of Use**

Of particular interest is the use of the nucleotide sequences in recombinant  
25 DNA constructs to direct the transcription or transcription and translation (expression) of the prenyltransferase or tocopherol cyclase sequences of the present invention in a host plant cell. The expression constructs generally comprise a promoter functional in a host plant cell operably linked to a nucleic acid sequence encoding a

prenyltransferase or tocopherol cyclase of the present invention and a transcriptional termination region functional in a host plant cell.

A first nucleic acid sequence is "operably linked" or "operably associated" with a second nucleic acid sequence when the sequences are so arranged that the first  
5 nucleic acid sequence affects the function of the second nucleic-acid sequence. Preferably, the two sequences are part of a single contiguous nucleic acid molecule and more preferably are adjacent. For example, a promoter is operably linked to a gene if the promoter regulates or mediates transcription of the gene in a cell.

Those skilled in the art will recognize that there are a number of promoters  
10 which are functional in plant cells, and have been described in the literature. Chloroplast and plastid specific promoters, chloroplast or plastid functional promoters, and chloroplast or plastid operable promoters are also envisioned.

One set of plant functional promoters are constitutive promoters such as the CaMV35S or FMV35S promoters that yield high levels of expression in most plant  
15 organs. Enhanced or duplicated versions of the CaMV35S and FMV35S promoters are useful in the practice of this invention (Odell, *et al.* (1985) *Nature* 313:810-812; Rogers, U.S. Patent Number 5,378, 619). In addition, it may also be preferred to bring about expression of the prenyltransferase or tocopherol cyclase gene in specific tissues of the plant, such as leaf, stem, root, tuber, seed, fruit, etc., and the promoter chosen  
20 should have the desired tissue and developmental specificity.

Of particular interest is the expression of the nucleic acid sequences of the present invention from transcription initiation regions which are preferentially expressed in a plant seed tissue. Examples of such seed preferential transcription initiation sequences include those sequences derived from sequences encoding plant  
25 storage protein genes or from genes involved in fatty acid biosynthesis in oilseeds. Examples of such promoters include the 5' regulatory regions from such genes as napin (Kridl *et al.*, *Seed Sci. Res.* 1:209:219 (1991)), phaseolin, zein, soybean trypsin inhibitor, ACP, stearyl-ACP desaturase, soybean  $\alpha'$  subunit of  $\beta$ -conglycinin (soy 7s, (Chen *et al.*, *Proc. Natl. Acad. Sci.*, 83:8560-8564 (1986))) and oleosin.



It may be advantageous to direct the localization of proteins conferring prenyltransferase or tocopherol cyclase to a particular subcellular compartment, for example, to the mitochondrion, endoplasmic reticulum, vacuoles, chloroplast or other plastidic compartment. For example, where the genes of interest of the present invention will be targeted to plastids, such as chloroplasts, for expression, the constructs will also employ the use of sequences to direct the gene to the plastid. Such sequences are referred to herein as chloroplast transit peptides (CTP) or plastid transit peptides (PTP). In this manner, where the gene of interest is not directly inserted into the plastid, the expression construct will additionally contain a gene encoding a transit peptide to direct the gene of interest to the plastid. The chloroplast transit peptides may be derived from the gene of interest, or may be derived from a heterologous sequence having a CTP. Such transit peptides are known in the art. See, for example, Von Heijne *et al.* (1991) *Plant Mol. Biol. Rep.* 9:104-126; Clark *et al.* (1989) *J. Biol. Chem.* 264:17544-17550; della-Cioppa *et al.* (1987) *Plant Physiol.* 84:965-968; Romer *et al.* (1993) *Biochem. Biophys. Res Commun.* 196:1414-1421; and, Shah *et al.* (1986) *Science* 233:478-481.

Depending upon the intended use, the constructs may contain the nucleic acid sequence which encodes the entire prenyltransferase or tocopherol cyclase protein, or a portion thereof. For example, where antisense inhibition of a given prenyltransferase or tocopherol cyclase protein is desired, the entire prenyltransferase or tocopherol cyclase sequence is not required. Furthermore, where prenyltransferase or tocopherol cyclase sequences used in constructs are intended for use as probes, it may be advantageous to prepare constructs containing only a particular portion of a prenyltransferase or tocopherol cyclase encoding sequence, for example a sequence which is discovered to encode a highly conserved prenyltransferase or tocopherol cyclase region.

The skilled artisan will recognize that there are various methods for the inhibition of expression of endogenous sequences in a host cell. Such methods include, but are not limited to, antisense suppression (Smith, *et al.* (1988) *Nature*



334:724-726), co-suppression (Napoli, *et al.* (1989) *Plant Cell* 2:279-289),  
ribozymes (PCT Publication WO 97/10328), and combinations of sense and antisense  
Waterhouse, *et al.* (1998) *Proc. Natl. Acad. Sci. USA* 95:13959-13964. Methods for  
the suppression of endogenous sequences in a host cell typically employ the  
5 transcription or transcription and translation of at least a portion of the sequence to be  
suppressed. Such sequences may be homologous to coding as well as non-coding  
regions of the endogenous sequence.

Regulatory transcript termination regions may be provided in plant expression  
constructs of this invention as well. Transcript termination regions may be provided  
10 by the DNA sequence encoding the prenyltransferase or tocopherol cyclase or a  
convenient transcription termination region derived from a different gene source, for  
example, the transcript termination region which is naturally associated with the  
transcript initiation region. The skilled artisan will recognize that any convenient  
transcript termination region which is capable of terminating transcription in a plant  
15 cell may be employed in the constructs of the present invention.

Alternatively, constructs may be prepared to direct the expression of the  
prenyltransferase or tocopherol cyclase sequences directly from the host plant cell  
plastid. Such constructs and methods are known in the art and are generally  
described, for example, in Svab, *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:8526-  
20 8530 and Svab and Maliga (1993) *Proc. Natl. Acad. Sci. USA* 90:913-917 and in U.S.  
Patent Number 5,693,507.

The prenyltransferase or tocopherol cyclase constructs of the present invention  
can be used in transformation methods with additional constructs providing for the  
expression of other nucleic acid sequences encoding proteins involved in the  
25 production of tocopherols, or tocopherol precursors such as homogentisic acid and/or  
phytylpyrophosphate. Nucleic acid sequences encoding proteins involved in the  
production of homogentisic acid are known in the art, and include but not are limited  
to, 4-hydroxyphenylpyruvate dioxygenase (HPPD, EC 1.13.11.27) described for  
example, by Garcia, *et al.* ((1999) *Plant Physiol.* 119(4):1507-1516), mono or

bifunctional *tyrA* (described for example by Xia, *et al.* (1992) *J. Gen Microbiol.* 138:1309-1316, and Hudson, *et al.* (1984) *J. Mol. Biol.* 180:1023-1051), Oxygenase, 4-hydroxyphenylpyruvate di- (9CI), 4-Hydroxyphenylpyruvate dioxygenase; p-Hydroxyphenylpyruvate dioxygenase; p-Hydroxyphenylpyruvate hydroxylase; 5 p-Hydroxyphenylpyruvate oxidase; p-Hydroxyphenylpyruvic acid hydroxylase; p-Hydroxyphenylpyruvic hydroxylase; p-Hydroxyphenylpyruvic oxidase), 4-hydroxyphenylacetate, NAD(P)H:oxygen oxidoreductase (1-hydroxylating); 4-hydroxyphenylacetate 1-monooxygenase, and the like. In addition, constructs for the expression of nucleic acid sequences encoding proteins involved in the production of phytylpyrophosphate can also be employed with the prenyltransferase or tocopherol 10 cyclase constructs of the present invention. Nucleic acid sequences encoding proteins involved in the production of phytylpyrophosphate are known in the art, and include, but are not limited to geranylgeranylpyrophosphate synthase (GGPPS), geranylgeranylpyrophosphate reductase (GGH), 1-deoxyxylulose-5-phosphate 15 synthase, 1- deoxy-D-xylulose-5-phosphate reductoisomerase, 4-diphosphocytidyl-2-C-methylerythritol synthase, isopentyl pyrophosphate isomerase.

The prenyltransferase or tocopherol cyclase sequences of the present invention find use in the preparation of transformation constructs having a second expression cassette for the expression of additional sequences involved in tocopherol 20 biosynthesis. Additional tocopherol biosynthesis sequences of interest in the present invention include, but are not limited to gamma-tocopherol methyltransferase (Shintani, *et al.* (1998) *Science* 282(5396):2098-2100), tocopherol cyclase, and tocopherol methyltransferase.

A plant cell, tissue, organ, or plant into which the recombinant DNA 25 constructs containing the expression constructs have been introduced is considered transformed, transfected, or transgenic. A transgenic or transformed cell or plant also includes progeny of the cell or plant and progeny produced from a breeding program employing such a transgenic plant as a parent in a cross and exhibiting an altered

phenotype resulting from the presence of a prenyltransferase or tocopherol cyclase nucleic acid sequence.

Plant expression or transcription constructs having a prenyltransferase or tocopherol cyclase as the DNA sequence of interest for increased or decreased  
5 expression thereof may be employed with a wide variety of plant life, particularly, plant life involved in the production of vegetable oils for edible and industrial uses. Particularly preferred plants for use in the methods of the present invention include, but are not limited to: *Acacia*, alfalfa, aneth, apple, apricot, artichoke, arugula, asparagus, avocado, banana, barley, beans, beet, blackberry, blueberry, broccoli,  
10 brussels sprouts, cabbage, canola, cantaloupe, carrot, cassava, cauliflower, celery, cherry, chicory, cilantro, citrus, clementines, coffee, corn, cotton, cucumber, Douglas fir, eggplant, endive, escarole, eucalyptus, fennel, figs, garlic, gourd, grape, grapefruit, honey dew, jicama, kiwifruit, lettuce, leeks, lemon, lime, Loblolly pine, mango, melon, mushroom, nectarine, nut, oat, oil palm, oil seed rape, okra, onion, orange, an  
15 ornamental plant, papaya, parsley, pea, peach, peanut, pear, pepper, persimmon, pine, pineapple, plantain, plum, pomegranate, poplar, potato, pumpkin, quince, radiata pine, radicchio, radish, raspberry, rice, rye, sorghum, Southern pine, soybean, spinach, squash, strawberry, sugarbeet, sugarcane, sunflower, sweet potato, sweetgum, tangerine, tea, tobacco, tomato, triticale, turf, turnip, a vine, watermelon, wheat, yams,  
20 and zucchini.

Most especially preferred are temperate oilseed crops. Temperate oilseed crops of interest include, but are not limited to, rapeseed (Canola and High Erucic Acid varieties), sunflower, safflower, cotton, soybean, peanut, coconut and oil palms, and corn. Depending on the method for introducing the recombinant constructs into  
25 the host cell, other DNA sequences may be required. Importantly, this invention is applicable to dicotyledons and monocotyledons species alike and will be readily applicable to new and/or improved transformation and regulation techniques.

Of particular interest, is the use of prenyltransferase or tocopherol cyclase constructs in plants to produce plants or plant parts, including, but not limited to

leaves, stems, roots, reproductive, and seed, with a modified content of tocopherols in plant parts having transformed plant cells.

For immunological screening, antibodies to the protein can be prepared by injecting rabbits or mice with the purified protein or portion thereof, such methods of preparing antibodies being well known to those in the art. Either monoclonal or polyclonal antibodies can be produced, although typically polyclonal antibodies are more useful for gene isolation. Western analysis may be conducted to determine that a related protein is present in a crude extract of the desired plant species, as determined by cross-reaction with the antibodies to the encoded proteins. When cross-reactivity is observed, genes encoding the related proteins are isolated by screening expression libraries representing the desired plant species. Expression libraries can be constructed in a variety of commercially available vectors, including lambda gt11, as described in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual*, Second Edition (1989) Cold Spring Harbor Laboratory, Cold Spring Harbor, New York).

To confirm the activity and specificity of the proteins encoded by the identified nucleic acid sequences as prenyltransferase or tocopherol cyclase enzymes, *in vitro* assays are performed in insect cell cultures using baculovirus expression systems. Such baculovirus expression systems are known in the art and are described by Lee, *et al.* U.S. Patent Number 5,348,886, the entirety of which is herein incorporated by reference.

In addition, other expression constructs may be prepared to assay for protein activity utilizing different expression systems. Such expression constructs are transformed into yeast or prokaryotic host and assayed for prenyltransferase or tocopherol cyclase activity. Such expression systems are known in the art and are readily available through commercial sources.

In addition to the sequences described in the present invention, DNA coding sequences useful in the present invention can be derived from algae, fungi, bacteria, mammalian sources, plants, etc. Homology searches in existing databases using

signature sequences corresponding to conserved nucleotide and amino acid sequences of prenyltransferase or tocopherol cyclase can be employed to isolate equivalent, related genes from other sources such as plants and microorganisms. Searches in EST databases can also be employed. Furthermore, the use of DNA sequences  
5 encoding enzymes functionally enzymatically equivalent to those disclosed herein, wherein such DNA sequences are degenerate equivalents of the nucleic acid sequences disclosed herein in accordance with the degeneracy of the genetic code, is also encompassed by the present invention. Demonstration of the functionality of coding sequences identified by any of these methods can be carried out by  
10 complementation of mutants of appropriate organisms, such as *Synechocystis*, *Shewanella*, yeast, *Pseudomonas*, *Rhodobacteria*, etc., that lack specific biochemical reactions, or that have been mutated. The sequences of the DNA coding regions can be optimized by gene resynthesis, based on codon usage, for maximum expression in particular hosts.

15 For the alteration of tocopherol production in a host cell, a second expression construct can be used in accordance with the present invention. For example, the prenyltransferase or tocopherol cyclase expression construct can be introduced into a host cell in conjunction with a second expression construct having a nucleotide sequence for a protein involved in tocopherol biosynthesis.

20 The method of transformation in obtaining such transgenic plants is not critical to the instant invention, and various methods of plant transformation are currently available. Furthermore, as newer methods become available to transform crops, they may also be directly applied hereunder. For example, many plant species naturally susceptible to *Agrobacterium* infection may be successfully transformed via tripartite  
25 or binary vector methods of *Agrobacterium* mediated transformation. In many instances, it will be desirable to have the construct bordered on one or both sides by T-DNA, particularly having the left and right borders, more particularly the right border. This is particularly useful when the construct uses *A. tumefaciens* or *A. rhizogenes* as a mode for transformation, although the T-DNA borders may find use

with other modes of transformation. In addition, techniques of microinjection, DNA particle bombardment, and electroporation have been developed which allow for the transformation of various monocot and dicot plant species.

Normally, included with the DNA construct will be a structural gene having  
5 the necessary regulatory regions for expression in a host and providing for selection of transformant cells. The gene may provide for resistance to a cytotoxic agent, e.g. antibiotic, heavy metal, toxin, etc., complementation providing prototrophy to an auxotrophic host, viral immunity or the like. Depending upon the number of different host species the expression construct or components thereof are introduced, one or  
10 more markers may be employed, where different conditions for selection are used for the different hosts.

Where *Agrobacterium* is used for plant cell transformation, a vector may be used which may be introduced into the *Agrobacterium* host for homologous recombination with T-DNA or the Ti- or Ri-plasmid present in the *Agrobacterium*  
15 host. The Ti- or Ri-plasmid containing the T-DNA for recombination may be armed (capable of causing gall formation) or disarmed (incapable of causing gall formation), the latter being permissible, so long as the *vir* genes are present in the transformed *Agrobacterium* host. The armed plasmid can give a mixture of normal plant cells and gall.

20 In some instances where *Agrobacterium* is used as the vehicle for transforming host plant cells, the expression or transcription construct bordered by the T-DNA border region(s) will be inserted into a broad host range vector capable of replication in *E. coli* and *Agrobacterium*, there being broad host range vectors described in the literature. Commonly used is pRK2 or derivatives thereof. See, for example, Ditta, *et*  
25 *al.*, (*Proc. Nat. Acad. Sci., U.S.A.* (1980) 77:7347-7351) and EPA 0 120 515, which are incorporated herein by reference. Alternatively, one may insert the sequences to be expressed in plant cells into a vector containing separate replication sequences, one of which stabilizes the vector in *E. coli*, and the other in *Agrobacterium*. See, for example, McBride, *et al.* (*Plant Mol. Biol.* (1990) 14:269-276), wherein the pRiHRI



(Jouanin, *et al.*, *Mol. Gen. Genet.* (1985) 201:370-374) origin of replication is utilized and provides for added stability of the plant expression vectors in host *Agrobacterium* cells.

Included with the expression construct and the T-DNA will be one or more  
5 markers, which allow for selection of transformed *Agrobacterium* and transformed plant cells. A

number of markers have been developed for use with plant cells, such as resistance to chloramphenicol, kanamycin, the aminoglycoside G418, hygromycin, or the like. The particular marker employed is not essential to this invention, one or another marker  
10 being preferred depending on the particular host and the manner of construction.

For transformation of plant cells using *Agrobacterium*, explants may be combined and incubated with the transformed *Agrobacterium* for sufficient time for transformation, the bacteria killed, and the plant cells cultured in an appropriate selective medium. Once callus forms, shoot formation can be encouraged by  
15 employing the appropriate plant hormones in accordance with known methods and the shoots transferred to rooting medium for regeneration of plants. The plants may then be grown to seed and the seed used to establish repetitive generations and for isolation of vegetable oils.

There are several possible ways to obtain the plant cells of this invention  
20 which contain multiple expression constructs. Any means for producing a plant comprising a construct having a DNA sequence encoding the expression construct of the present invention, and at least one other construct having another DNA sequence encoding an enzyme are encompassed by the present invention. For example, the expression construct can be used to transform a plant at the same time as the second  
25 construct either by inclusion of both expression constructs in a single transformation vector or by using separate vectors, each of which express desired genes. The second construct can be introduced into a plant which has already been transformed with the prenyltransferase or tocopherol cyclase expression construct, or alternatively, transformed plants, one expressing the prenyltransferase or tocopherol cyclase



construct and one expressing the second construct, can be crossed to bring the constructs together in the same plant.

Transgenic plants of the present invention may be produced from tissue culture, and subsequent generations grown from seed. Alternatively, transgenic plants  
5 may be grown using apomixis. Apomixis is a genetically controlled method of reproduction in plants where the embryo is formed without union of an egg and a sperm. There are three basic types of apomictic reproduction: 1) apospory where the embryo develops from a chromosomally unreduced egg in an embryo sac derived from the nucleus, 2) diplospory where the embryo develops from an unreduced egg in  
10 an embryo sac derived from the megaspore mother cell, and 3) adventitious embryony where the embryo develops directly from a somatic cell. In most forms of apomixis, pseudogamy or fertilization of the polar nuclei to produce endosperm is necessary for seed viability. In apospory, a nurse cultivar can be used as a pollen source for endosperm formation in seeds. The nurse cultivar does not affect the genetics of the  
15 aposporous apomictic cultivar since the unreduced egg of the cultivar develops parthenogenetically, but makes possible endosperm production. Apomixis is economically important, especially in transgenic plants, because it causes any genotype, no matter how heterozygous, to breed true. Thus, with apomictic reproduction, heterozygous transgenic plants can maintain their genetic fidelity  
20 throughout repeated life cycles. Methods for the production of apomictic plants are known in the art. See, U.S. Patent No. 5,811,636, which is herein incorporated by reference in its entirety.

The nucleic acid sequences of the present invention can be used in constructs to provide for the expression of the sequence in a variety of host cells, both  
25 prokaryotic eukaryotic. Host cells of the present invention preferably include monocotyledenous and dicotyledenous plant cells.

In general, the skilled artisan is familiar with the standard resource materials which describe specific conditions and procedures for the construction, manipulation and isolation of macromolecules (e.g., DNA molecules, plasmids, etc.), generation of

recombinant organisms and the screening and isolating of clones, (see for example, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press (1989); Maliga *et al.*, *Methods in Plant Molecular Biology*, Cold Spring Harbor Press (1995), the entirety of which is herein incorporated by reference; Birren *et al.*,  
5 *Genome Analysis: Analyzing DNA*, 1, Cold Spring Harbor, New York, the entirety of which is herein incorporated by reference).

Methods for the expression of sequences in insect host cells are known in the art. Baculovirus expression vectors are recombinant insect viruses in which the coding sequence for a chosen foreign gene has been inserted behind a baculovirus promoter  
10 in place of the viral gene, e.g., polyhedrin (Smith and Summers, U.S. Pat. No., 4,745,051, the entirety of which is incorporated herein by reference). Baculovirus expression vectors are known in the art, and are described for example in Doerfler, *Curr. Top. Microbiol. Immunol.* 131:51-68 (1968); Luckow and Summers, *Bio/Technology* 6:47-55 (1988a); Miller, *Annual Review of Microbiol.* 42:177-199  
15 (1988); Summers, *Curr. Comm. Molecular Biology*, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1988); Summers and Smith, *A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures*, Texas Ag. Exper. Station Bulletin No. 1555 (1988), the entireties of which is herein incorporated by reference)

Methods for the expression of a nucleic acid sequence of interest in a fungal  
20 host cell are known in the art. The fungal host cell may, for example, be a yeast cell or a filamentous fungal cell. Methods for the expression of DNA sequences of interest in yeast cells are generally described in "Guide to yeast genetics and molecular biology", Guthrie and Fink, eds. *Methods in enzymology*, Academic Press, Inc. Vol 194 (1991) and *Gene expression technology*", Goeddel ed, *Methods in Enzymology*, Academic  
25 Press, Inc., Vol 185 (1991).

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC, Manassas, VA), such as HeLa cells, Chinese hamster ovary (CHO) cells, baby hamster kidney (BHK) cells and a number of other cell lines.

Suitable promoters for mammalian cells are also known in the art and include, but are not limited to, viral promoters such as that from Simian Virus 40 (SV40) (Fiers *et al.*, *Nature* 273:113 (1978), the entirety of which is herein incorporated by reference), Rous sarcoma virus (RSV), adenovirus (ADV) and bovine papilloma virus (BPV).

5 Mammalian cells may also require terminator sequences and poly-A addition sequences. Enhancer sequences which increase expression may also be included and sequences which promote amplification of the gene may also be desirable (for example methotrexate resistance genes).

Vectors suitable for replication in mammalian cells are well known in the art,

10 and may include viral replicons, or sequences which insure integration of the appropriate sequences encoding epitopes into the host genome. Plasmid vectors that greatly facilitate the construction of recombinant viruses have been described (*see*, for example, Mackett *et al.*, *J Virol.* 49:857 (1984); Chakrabarti *et al.*, *Mol. Cell. Biol.* 5:3403 (1985); Moss, In: *Gene Transfer Vectors For Mammalian Cells* (Miller and

15 Calos, eds., Cold Spring Harbor Laboratory, N.Y., p. 10, (1987); all of which are herein incorporated by reference in their entirety).

The invention also includes plants and plant parts, such as seed, oil and meal derived from seed, and feed and food products processed from plants, which are enriched in tocopherols. Of particular interest is seed oil obtained from transgenic

20 plants where the tocopherol level has been increased as compared to seed oil of a non-transgenic plant.

The harvested plant material may be subjected to additional processing to further enrich the tocopherol content. The skilled artisan will recognize that there are many such processes or methods for refining, bleaching and degumming oil. United

25 States Patent Number 5,932,261, issued August 3, 1999, discloses on such process, for the production of a natural carotene rich refined and deodorised oil by subjecting the oil to a pressure of less than 0.060 mbar and to a temperature of less than 200.degree. C. Oil distilled by this process has reduced free fatty acids, yielding a refined, deodorised oil where Vitamin E contained in the feed oil is substantially

retained in the processed oil. The teachings of this patent are incorporated herein by reference.

The invention now being generally described, it will be more readily  
5 understood by reference to the following examples which are included for purposes of illustration only and are not intended to limit the present invention.

### EXAMPLES

#### Example 1: Identification of Prenyltransferase or tocopherol cyclase Sequences

10 PSI-BLAST (Altschul, *et al.* (1997) *Nuc Acid Res* 25:3389-3402) profiles were generated for both the straight chain and aromatic classes of prenyltransferases. To generate the straight chain profile, a prenyl-transferase from *Porphyra purpurea* (Genbank accession 1709766) was used as a query against the NCBI non-redundant protein database. The *E. coli* enzyme involved in the formation of ubiquinone, ubiA  
15 (genbank accession 1790473) was used as a starting sequence to generate the aromatic prenyltransferase profile. These profiles were used to search public and proprietary DNA and protein data bases. In *Arabidopsis* six putative prenyltransferases of the straight-chain class were identified, ATPT1, (SEQ ID NO:9), ATPT7 (SEQ ID NO:10), ATPT8 (SEQ ID NO:11), ATPT9 (SEQ ID NO:13), ATPT10 (SEQ ID  
20 NO:14), and ATPT11 (SEQ ID NO:15), and six were identified of the aromatic class, ATPT2 (SEQ ID NO:1), ATPT3 (SEQ ID NO:3), ATPT4 (SEQ ID NO:5), ATPT5 (SEQ ID NO:7), ATPT6 (SEQ ID NO:8), and ATPT12 (SEQ ID NO:16). Additional prenyltransferase sequences from other plants related to the aromatic class of prenyltransferases, such as soy (SEQ ID NOs: 19-23, the deduced amino acid  
25 sequence of SEQ ID NO:23 is provided in SEQ ID NO:24) and maize (SEQ ID NOs:25-29, and 31) are also identified. The deduced amino acid sequence of ZMPT5 (SEQ ID NO:29) is provided in SEQ ID NO:30.

Searches are performed on a Silicon Graphics Unix computer using additional Bioaccelerator hardware and GenWeb software supplied by Compugen Ltd. This

software and hardware enables the use of the Smith-Waterman algorithm in searching DNA and protein databases using profiles as queries. The program used to query protein databases is profilesearch. This is a search where the query is not a single sequence but a profile based on a multiple alignment of amino acid or nucleic acid sequences. The profile is used to query a sequence data set, i.e., a sequence database. The profile contains all the pertinent information for scoring each position in a sequence, in effect replacing the "scoring matrix" used for the standard query searches. The program used to query nucleotide databases with a protein profile is tprofilesearch. Tprofilesearch searches nucleic acid databases using an amino acid profile query. As the search is running, sequences in the database are translated to amino acid sequences in six reading frames. The output file for tprofilesearch is identical to the output file for profilesearch except for an additional column that indicates the frame in which the best alignment occurred.

The Smith-Waterman algorithm, (Smith and Waterman (1981) *supra*), is used to search for similarities between one sequence from the query and a group of sequences contained in the database. E score values as well as other sequence information, such as conserved peptide sequences are used to identify related sequences.

To obtain the entire coding region corresponding to the *Arabidopsis* prenyltransferase sequences, synthetic oligo-nucleotide primers are designed to amplify the 5' and 3' ends of partial cDNA clones containing prenyltransferase sequences. Primers are designed according to the respective *Arabidopsis* prenyltransferase sequences and used in Rapid Amplification of cDNA Ends (RACE) reactions (Frohman *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:8998-9002) using the Marathon cDNA amplification kit (Clontech Laboratories Inc, Palo Alto, CA).

Amino acid sequence alignments between ATPT2 (SEQ ID NO:2), ATPT3 (SEQ ID NO:4), ATPT4 (SEQ ID NO:6), ATPT8 (SEQ ID NO:12), and ATPT12 (SEQ ID NO:17) are performed using ClustalW (Figure 1), and the percent identity and similarities are provided in Table 1 below.

**Table 1:**

	ATPT2	ATPT3	ATPT4	ATPT8	ATPT12
ATPT2 % Identity	12	13	11	15	
% similar	25	25	22	32	
% Gap	17	20	20	9	
ATPT3 % Identity		12	6	22	
% similar		29	16	38	
% Gap		20	24	14	
ATPT4 % Identity			9	14	
% similar			18	29	
% Gap			26	19	
ATPT8 % Identity				7	
% similar				19	
% Gap				20	
ATPT12 % Identity					
% similar					
% Gap					

**Example 2: Preparation of Prenyl Transferase Expression Constructs**

5 A plasmid containing the napin cassette derived from pCGN3223 (described in USPN 5,639,790, the entirety of which is incorporated herein by reference) was modified to make it more useful for cloning large DNA fragments containing multiple restriction sites, and to allow the cloning of multiple napin fusion genes into plant binary transformation vectors. An adapter comprised of the self annealed

10 oligonucleotide of sequence

CGCGATTAAATGGCGCGCCCTGCAGGCGGCCGCTGCAGGGCGCGCCAT  
TAAAT (SEQ ID NO:40) was ligated into the cloning vector pBC SK+ (Stratagene) after digestion with the restriction endonuclease BssHII to construct vector

pCGN7765. Plasmids pCGN3223 and pCGN7765 were digested with NotI and ligated together. The resultant vector, pCGN7770, contains the pCGN7765 backbone with the napin seed specific expression cassette from pCGN3223.

The cloning cassette, pCGN7787, essentially the same regulatory elements as pCGN7770, with the exception of the napin regulatory regions of pCGN7770 have been replaced with the double CAMV 35S promoter and the tml polyadenylation and transcriptional termination region.

A binary vector for plant transformation, pCGN5139, was constructed from pCGN1558 (McBride and Summerfelt, (1990) Plant Molecular Biology, 14:269-276). The polylinker of pCGN1558 was replaced as a HindIII/Asp718 fragment with a polylinker containing unique restriction endonuclease sites, AscI, PacI, XbaI, SmaI, BamHI, and NotI. The Asp718 and HindIII restriction endonuclease sites are retained in pCGN5139.

A series of turbo binary vectors are constructed to allow for the rapid cloning of DNA sequences into binary vectors containing transcriptional initiation regions (promoters) and transcriptional termination regions.

The plasmid pCGN8618 was constructed by ligating oligonucleotides 5'-TCGAGGATCCGCGGCCGCAAGCTTCCTGCAGG-3' (SEQ ID NO:41) and 5'-TCGACCTGCAGGAAGCTTGCGGCCGCGGATCC-3' (SEQ ID NO:42) into Sall/XhoI-digested pCGN7770. A fragment containing the napin promoter, polylinker and napin 3' region was excised from pCGN8618 by digestion with Asp718I; the fragment was blunt-ended by filling in the 5' overhangs with Klenow fragment then ligated into pCGN5139 that had been digested with Asp718I and HindIII and blunt-ended by filling in the 5' overhangs with Klenow fragment. A plasmid containing the insert oriented so that the napin promoter was closest to the blunted Asp718I site of pCGN5139 and the napin 3' was closest to the blunted HindIII site was subjected to sequence analysis to confirm both the insert orientation and the integrity of cloning junctions. The resulting plasmid was designated pCGN8622.



The plasmid pCGN8619 was constructed by ligating oligonucleotides 5'-TCGACCTGCAGGAAGCTTGCGGCCGCGGATCC -3' (SEQ ID NO:43) and 5'-TCGAGGATCCGCGGCCGCAAGCTTCCTGCAGG-3' (SEQ ID NO:44) into SalI/XhoI-digested pCGN7770. A fragment containing the napin promoter,  
5 polylinker and napin 3' region was removed from pCGN8619 by digestion with Asp718I; the fragment was blunt-ended by filling in the 5' overhangs with Klenow fragment then ligated into pCGN5139 that had been digested with Asp718I and HindIII and blunt-ended by filling in the 5' overhangs with Klenow fragment. A plasmid containing the insert oriented so that the napin promoter was closest to the  
10 blunted Asp718I site of pCGN5139 and the napin 3' was closest to the blunted HindIII site was subjected to sequence analysis to confirm both the insert orientation and the integrity of cloning junctions. The resulting plasmid was designated pCGN8623.

The plasmid pCGN8620 was constructed by ligating oligonucleotides 5'-  
15 TCGAGGATCCGCGGCCGCAAGCTTCCTGCAGGAGCT -3' (SEQ ID NO:45) and 5'-CCTGCAGGAAGCTTGCGGCCGCGGATCC-3' (SEQ ID NO:46) into SalI/SacI-digested pCGN7787. A fragment containing the d35S promoter, polylinker and tml 3' region was removed from pCGN8620 by complete digestion with Asp718I and partial digestion with NotI. The fragment was blunt-ended by filling in the 5'  
20 overhangs with Klenow fragment then ligated into pCGN5139 that had been digested with Asp718I and HindIII and blunt-ended by filling in the 5' overhangs with Klenow fragment. A plasmid containing the insert oriented so that the d35S promoter was closest to the blunted Asp718I site of pCGN5139 and the tml 3' was closest to the blunted HindIII site was subjected to sequence analysis to confirm both the insert  
25 orientation and the integrity of cloning junctions. The resulting plasmid was designated pCGN8624.

The plasmid pCGN8621 was constructed by ligating oligonucleotides 5'-TCGACCTGCAGGAAGCTTGCGGCCGCGGATCCAGCT -3' (SEQ ID NO:47) and 5'-GGATCCGCGGCCGCAAGCTTCCTGCAGG-3' (SEQ ID NO:48) into

Sall/SacI-digested pCGN7787. A fragment containing the d35S promoter, polylinker and tml 3' region was removed from pCGN8621 by complete digestion with Asp718I and partial digestion with NotI. The fragment was blunt-ended by filling in the 5' overhangs with Klenow fragment then ligated into pCGN5139 that had been digested with Asp718I and HindIII and blunt-ended by filling in the 5' overhangs with Klenow fragment. A plasmid containing the insert oriented so that the d35S promoter was closest to the blunted Asp718I site of pCGN5139 and the tml 3' was closest to the blunted HindIII site was subjected to sequence analysis to confirm both the insert orientation and the integrity of cloning junctions. The resulting plasmid was designated pCGN8625.

The plasmid construct pCGN8640 is a modification of pCGN8624 described above. A 938bp PstI fragment isolated from transposon Tn7 which encodes bacterial spectinomycin and streptomycin resistance (Fling et al. (1985), *Nucleic Acids Research* 13(19):7095-7106), a determinant for *E. coli* and *Agrobacterium* selection, was blunt ended with Pfu polymerase. The blunt ended fragment was ligated into pCGN8624 that had been digested with SpeI and blunt ended with Pfu polymerase. The region containing the PstI fragment was sequenced to confirm both the insert orientation and the integrity of cloning junctions.

The spectinomycin resistance marker was introduced into pCGN8622 and pCGN8623 as follows. A 7.7 Kbp AvrII-SnaBI fragment from pCGN8640 was ligated to a 10.9 Kbp AvrII-SnaBI fragment from pCGN8623 or pCGN8622, described above. The resulting plasmids were pCGN8641 and pCGN8643, respectively.

The plasmid pCGN8644 was constructed by ligating oligonucleotides 5'-GATCACCTGCAGGAAGCTTGCGGCCGCGGATCCAATGCA-3' (SEQ ID NO:49) and 5'-TTGGATCCGCGGCCGCAAGCTTCCTGCAGGT-3' (SEQ ID NO:50) into BamHI-PstI digested pCGN8640.

Synthetic oligonucleotides were designed for use in Polymerase Chain Reactions (PCR) to amplify the coding sequences of ATPT2, ATPT3, ATPT4, ATPT8, and ATPT12 for the preparation of expression constructs and are provided in Table 2 below.

5 **Table 2:**

Name	Restriction Site	Sequence	SEQ ID NO:
ATPT2	5' NotI	GGATCCGCGGCCGCACAATGGAGTC TCTGCTCTCTAGTTCT	51
ATPT2	3' SseI	GGATCCTGCAGGTCACCTCAAAAAA GGTAACAGCAAGT	52
ATPT3	5' NotI	GGATCCGCGGCCGCACAATGGCGTT TTTTGGGCTCTCCCGTGTTT	53
ATPT3	3' SseI	GGATCCTGCAGGTTATTGAAAACCTT CTTCCAAGTACAACT	54
ATPT4	5' NotI	GGATCCGCGGCCGCACAATGTGGCG AAGATCTGTTGTT	55
ATPT4	3' SseI	GGATCCTGCAGGTCATGGAGAGTAG AAGGAAGGAGCT	56
ATPT8	5' NotI	GGATCCGCGGCCGCACAATGGTACT TGCCGAGGTTCCAAAGCTTGCCTCT	57
ATPT8	3' SseI	GGATCCTGCAGGTCACCTTGTTTCTGG TGATGACTCTAT	58
ATPT12	5' NotI	GGATCCGCGGCCGCACAATGACTTC GATTCTCAACACT	59
ATPT12	3' SseI	GGATCCTGCAGGTCAGTGTTGCGAT GCTAATGCCGT	60

The coding sequences of ATPT2, ATPT3, ATPT4, ATPT8, and ATPT12 were all amplified using the respective PCR primers shown in Table 2 above and cloned into the TopoTA vector (Invitrogen). Constructs containing the respective prenyltransferase sequences were digested with NotI and Sse8387I and cloned into the turbobinary vectors described above.

The sequence encoding ATPT2 prenyltransferase was cloned in the sense orientation into pCGN8640 to produce the plant transformation construct pCGN10800 (Figure 2). The ATPT2 sequence is under control of the 35S promoter.

The ATPT2 sequence was also cloned in the antisense orientation into the construct pCGN8641 to create pCGN10801 (Figure 3). This construct provides for the antisense expression of the ATPT2 sequence from the napin promoter.

The ATPT2 coding sequence was also cloned in the sense orientation into the  
5 vector pCGN8643 to create the plant transformation construct pCGN10822

The ATPT2 coding sequence was also cloned in the antisense orientation into the vector pCGN8644 to create the plant transformation construct pCGN10803 (Figure 4).

The ATPT4 coding sequence was cloned into the vector pCGN864 to create the plant transformation construct pCGN10806 (Figure 5). The ATPT2 coding sequence was  
10 cloned into the vector TopoTA™ vector from Invitrogen, to create the plant transformation construct pCGN10807 (Figure 6). The ATPT3 coding sequence was cloned into the TopoTA vector to create the plant transformation construct pCGN10808 (Figure 7). The ATPT3 coding sequence was cloned in the sense orientation into the vector pCGN8640 to create the plant transformation construct pCGN10809 (Figure 8). The  
15 ATPT3 coding sequence was cloned in the antisense orientation into the vector pCGN8641 to create the plant transformation construct pCGN10810 (Figure 9). The ATPT3 coding sequence was cloned into the vector pCGN8643 to create the plant transformation construct pCGN10811 (Figure 10). The ATPT3 coding sequence was cloned into the vector pCGN8644 to create the plant transformation construct  
20 pCGN10812 (Figure 11). The ATPT4 coding sequence was cloned into the vector pCGN8640 to create the plant transformation construct pCGN10813 (Figure 12). The ATPT4 coding sequence was cloned into the vector pCGN8641 to create the plant transformation construct pCGN10814 (Figure 13). The ATPT4 coding sequence was cloned into the vector pCGN8643 to create the plant transformation construct  
25 pCGN10815 (Figure 14). The ATPT4 coding sequence was cloned in the antisense orientation into the vector pCGN8644 to create the plant transformation construct pCGN10816 (Figure 15). The ATPT8 coding sequence was cloned in the sense orientation into the vector pCGN8643 to create the plant transformation construct pCGN10819 (Figure 17). The ATPT12 coding sequence was cloned into the vector

pCGN8640 to create the plant transformation construct pCGN10824 (Figure 18). The ATPT12 coding sequence was cloned into the vector pCGN8643 to create the plant transformation construct pCGN10825 (Figure 19). The ATPT8 coding sequence was cloned into the vector pCGN8640 to create the plant transformation construct  
5 pCGN10826 (Figure 20).

### Example 3: Plant Transformation with Prenyl Transferase Constructs

Transgenic *Brassica* plants are obtained by *Agrobacterium*-mediated transformation as described by Radke *et al.* (*Theor. Appl. Genet.* (1988) 75:685-694;  
10 *Plant Cell Reports* (1992) 11:499-505). Transgenic *Arabidopsis thaliana* plants may be obtained by *Agrobacterium*-mediated transformation as described by Valverkens *et al.*, (*Proc. Nat. Acad. Sci.* (1988) 85:5536-5540), or as described by Bent *et al.* ((1994), *Science* 265:1856-1860), or Bechtold *et al.* ((1993), *C.R.Acad.Sci, Life Sciences* 316:1194-1199). Other plant species may be similarly transformed using  
15 related techniques.

Alternatively, microprojectile bombardment methods, such as described by Klein *et al.* (*Bio/Technology* 10:286-291) may also be used to obtain nuclear transformed plants.

### 20 Example 4: Identification of Additional Prenyltransferases

Additional BLAST searches were performed using the ATPT2 sequence, a sequence in the class of aromatic prenyltransferases. ESTs, and in some case, full-length coding regions, were identified in proprietary DNA libraries.

Soy full-length homologs to ATPT2 were identified by a combination of  
25 BLAST (using ATPT2 protein sequence) and 5' RACE. Two homologs resulted (SEQ ID NO:95 and SEQ ID NO:96). Translated amino acid sequences are provided by SEQ ID NO:97 and SEQ ID NO:98.

A rice est ATPT2 homolog is shown in SEQ ID NO:99 (obtained from BLAST using the wheat ATPT2 homolog).

Other homolog sequences were obtained using ATPT2 and PSI-BLAST, including est sequences from wheat (SEQ ID NO:100), leek (SEQ ID NOs:101 and 102), canola (SEQ ID NO:103), corn (SEQ ID NOs:104, 105 and 106), cotton (SEQ ID NO:107) and tomato (SEQ ID NO:108).

5 A PSI-Blast profile generated using the *E. coli* ubiA (genbank accession 1790473) sequence was used to analyze the *Synechocystis* genome. This analysis identified 5 open reading frames (ORFs) in the *Synechocystis* genome that were potentially prenyltransferases; slr0926 (annotated as ubiA (4-hydroxybenzoate-octaprenyltransferase, SEQ ID NO:32), slr1899 (annotated as ctaB (cytochrome c  
10 oxidase folding protein, SEQ ID NO:33), slr0056 (annotated as g4 (chlorophyll synthase 33 kd subunit, SEQ ID NO:34), slr1518 (annotated as menA (menaquinone biosynthesis protein, SEQ ID NO:35), and slr1736 (annotated as a hypothetical protein of unknown function (SEQ ID NO:36).

#### 15 4A. *Synechocystis* Knock-outs

To determine the functionality of these ORFs and their involvement, if any, in the biosynthesis of tocopherols, knockouts constructs were made to disrupt the ORF identified in *Synechocystis*.

Synthetic oligos were designed to amplify regions from the 5' (5'-  
20 TAATGTGTACATTGTCGGCCTC (17365') (SEQ ID NO:61) and 5'-  
GCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTCCACAATTCC  
CCGCACCGTC (1736kanpr1)) (SEQ ID NO:62) and 3' (5'-  
AGGCTAATAAGCACAAATGGGA (17363') (SEQ ID NO:63) and 5'-  
GGTATGAGTCAGCAACACCTTCTTCACGAGGCAGACCTCAGC  
25 GGAATTGGTTTAGGTTATCCC (1736kanpr2)) (SEQ ID NO:64) ends of the  
slr1736 ORF. The 1736kanpr1 and 1736kanpr2 oligos contained 20 bp of homology to the slr1736 ORF with an additional 40 bp of sequence homology to the ends of the kanamycin resistance cassette. Separate PCR steps were completed with these oligos and the products were gel purified and combined with the kanamycin resistance gene

from puc4K (Pharmacia) that had been digested with *HincII* and gel purified away from the vector backbone. The combined fragments were allowed to assemble without oligos under the following conditions: 94°C for 1 min, 55°C for 1 min, 72°C for 1 min plus 5 seconds per cycle for 40 cycles using pfu polymerase in 100ul reaction volume (Zhao, H and Arnold (1997) *Nucleic Acids Res.* 25(6):1307-1308). One microliter or five microliters of this assembly reaction was then amplified using 5' and 3' oligos nested within the ends of the ORF fragment, so that the resulting product contained 100-200 bp of the 5' end of the *Synechocystis* gene to be knocked out, the kanamycin resistance cassette, and 100-200 bp of the 3' end of the gene to be knocked out. This PCR product was then cloned into the vector pGemT easy (Promega) to create the construct pMON21681 and used for *Synechocystis* transformation.

Primers were also synthesized for the preparation of *Synechocystis* knockout constructs for the other sequences using the same method as described above, with the following primers. The *ubiA* 5' sequence was amplified using the primers 5'-GGATCCATGGTT GCCCAAACCCCATC (SEQ ID NO:65) and 5'-GCAATGTAACATCAGAGA TTTTGAGACACAACG TGGCTTTGGGTAAGCAACAATGACCGGC (SEQ ID NO:66). The 3' region was amplified using the synthetic oligonucleotide primers 5'-GAATTCTCAAAGCCAGCCCAGTAAC (SEQ ID NO:67) and 5'-GGTATGAGTC AGCAACACCTTCTTCACGAGGCAGACCTCAGCGGGTGCGAAAAGGGTTTT CCC (SEQ ID NO:68). The amplification products were combined with the kanamycin resistance gene from puc4K (Pharmacia) that had been digested with *HincII* and gel purified away from the vector backbone. The annealed fragment was amplified using 5' and 3' oligos nested within the ends of the ORF fragment (5'-CCAGTGGTTTAGGCTGTGTGGTC (SEQ ID NO:69) and 5'-CTGAGTTGGATGTATTGGATC (SEQ ID NO:70)), so that the resulting product contained 100-200 bp of the 5' end of the *Synechocystis* gene to be knocked out, the kanamycin resistance cassette, and 100-200 bp of the 3' end of the gene to be knocked



out. This PCR product was then cloned into the vector pGemT easy (Promega) to create the construct pMON21682 and used for *Synechocystis* transformation.

Primers were also synthesized for the preparation of *Synechocystis* knockout constructs for the other sequences using the same method as described above, with the following primers. The sl11899 5' sequence was amplified using the primers 5'-  
5 GGATCCATGGTTACTT CGACAAAAATCC (SEQ ID NO:71) and 5'-  
GCAATGTAACATCAGAG  
ATTTTGAGACACAACGTGGCTTTGCTAGGCAACCGCTTAGTAC (SEQ ID  
NO:72). The 3' region was amplified using the synthetic oligonucleotide primers 5'-  
10 GAATTCTTAACCCAACAGTAAAGTTCCC (SEQ ID NO:73) and 5'-  
GGTATGAGTCAGC  
AACACCTTCTTCACGAGGCAGACCTCAGCGCCGGCATTGTCTTTTACATG  
(SEQ ID NO:74). The amplification products were combined with the kanamycin  
resistance gene from puc4K (Pharmacia) that had been digested with *HincII* and gel  
15 purified away from the vector backbone. The annealed fragment was amplified using  
5' and 3' oligos nested within the ends of the ORF fragment (5'-  
GGAACCCTTGCAGCCGCTTC (SEQ ID NO:75)  
and 5'- GTATGCCCAACTGGTGCAGAGG (SEQ ID NO:76)), so that the resulting  
product contained 100-200 bp of the 5' end of the *Synechocystis* gene to be knocked  
20 out, the kanamycin resistance cassette, and 100-200 bp of the 3' end of the gene to be  
knocked out. This PCR product was then cloned into the vector pGemT easy  
(Promega) to create the construct pMON21679 and used for *Synechocystis*  
transformation.

Primers were also synthesized for the preparation of *Synechocystis* knockout  
25 constructs for the other sequences using the same method as described above, with the  
following primers. The slr0056 5' sequence was amplified using the primers 5'-  
GGATCCATGTCTGACACACAAAATACCG (SEQ ID NO:77) and 5'-  
GCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTTCGCCAATACC  
AGCCACCAACAG (SEQ ID NO:78). The 3' region was amplified using the

synthetic oligonucleotide primers 5'- GAATTCTCAAAT CCCC GCATGGCCTAG (SEQ ID NO:79) and 5'-  
GGTATGAGTCAGCAACACCTTCTTCACGAGGCAGACCTCAGCGGCCTACG  
GCTTGGACGTGTGGG (SEQ ID NO:80). The amplification products were  
5 combined with the kanamycin resistance gene from puc4K (Pharmacia) that had been  
digested with *HincII* and gel purified away from the vector backbone. The annealed  
fragment was amplified using 5' and 3' oligos nested within the ends of the ORF  
fragment (5'- CACTTGGATTCCCCTGATCTG (SEQ ID NO:81) and 5'-  
GCAATACCCGCTTGGAAAACG (SEQ ID NO:82)), so that the resulting product  
10 contained 100-200 bp of the 5' end of the *Synechocystis* gene to be knocked out, the  
kanamycin resistance cassette, and 100-200 bp of the 3' end of the gene to be knocked  
out. This PCR product was then cloned into the vector pGemT easy (Promega) to  
create the construct pMON21677 and used for *Synechocystis* transformation.

Primers were also synthesized for the preparation of *Synechocystis* knockout  
15 constructs for the other sequences using the same method as described above, with the  
following primers. The slr1518 5' sequence was amplified using the primers 5'-  
GGATCCATGACCGAAT CTTCGCCCCTAGC (SEQ ID NO:83) and 5'-  
GCAATGTAACATCAGAGATTTTGA GACACAACGTGGC  
TTTCAATCCTAGGTAGCCGAGGCG (SEQ ID NO:84). The 3' region was  
20 amplified using the synthetic oligonucleotide primers 5'-  
GAATTCTTAGCCCAGGCC AGCCCAGCC (SEQ ID NO:85) and 5'-  
GGTATGAGTCAGCAACACCTTCTTCACGA  
GGCAGACCTCAGCGGGGAATTGATTTGTTTAATTACC (SEQ ID NO:86). The  
amplification products were combined with the kanamycin resistance gene from  
25 puc4K (Pharmacia) that had been digested with *HincII* and gel purified away from the  
vector backbone. The annealed fragment was amplified using 5' and 3' oligos nested  
within the ends of the ORF fragment (5'- GCGATCGCCATTATCGCTTGG (SEQ ID  
NO:87) and 5'- GCAGACTGGCAATTATCAGTAACG (SEQ ID NO:88)), so that  
the resulting product contained 100-200 bp of the 5' end of the *Synechocystis* gene to

be knocked out, the kanamycin resistance cassette, and 100-200 bp of the 3' end of the gene to be knocked out. This PCR product was then cloned into the vector pGemT easy (Promega) to create the construct pMON21680 and used for *Synechocystis* transformation.

5

#### 4B. Transformation of *Synechocystis*

Cells of *Synechocystis* 6803 were grown to a density of approximately  $2 \times 10^8$  cells per ml and harvested by centrifugation. The cell pellet was re-suspended in fresh BG-11 medium (ATCC Medium 616) at a density of  $1 \times 10^9$  cells per ml and used immediately for transformation. One-hundred microliters of these cells were mixed with 5 ul of mini prep DNA and incubated with light at 30C for 4 hours. This mixture was then plated onto nylon filters resting on BG-11 agar supplemented with TES pH8 and allowed to grow for 12-18 hours. The filters were then transferred to BG-11 agar + TES + 5ug/ml kanamycin and allowed to grow until colonies appeared within 7-10 days (Packer and Glazer, 1988). Colonies were then picked into BG-11 liquid media containing 5 ug/ml kanamycin and allowed to grow for 5 days. These cells were then transferred to Bg-11 media containing 10ug/ml kanamycin and allowed to grow for 5 days and then transferred to Bg-11 + kanamycin at 25ug/ml and allowed to grow for 5 days. Cells were then harvested for PCR analysis to determine the presence of a disrupted ORF and also for HPLC analysis to determine if the disruption had any effect on tocopherol levels.

PCR analysis of the *Synechocystis* isolates for slr1736 and sl1899 showed complete segregation of the mutant genome, meaning no copies of the wild type genome could be detected in these strains. This suggests that function of the native gene is not essential for cell function. HPLC analysis of these same isolates showed that the sl1899 strain had no detectable reduction in tocopherol levels. However, the strain carrying the knockout for slr1736 produced no detectable levels of tocopherol.

The amino acid sequences for the *Synechocystis* knockouts are compared using ClustalW, and are provided in Table 3 below. Provided are the percent identities,

percent similarity, and the percent gap. The alignment of the sequences is provided in Figure 21.

**Table 3:**

	Slr1736	slr0926	sll1899	slr0056	slr1518
slr1736 %identity	14	12	18	11	
%similar	29	30	34	26	
%gap	8	7	10	5	
slr0926 %identity		20	19	14	
%similar		39	32	28	
%gap		7	9	4	
sll1899 %identity			17	13	
%similar			29	29	
%gap			12	9	
slr0056 %identity				15	
%similar				31	
%gap				8	
slr1518 %identity					
%similar					
%gap					

5

Amino acid sequence comparisons are performed using various *Arabidopsis* prenyltransferase sequences and the *Synechocystis* sequences. The comparisons are presented in Table 4 below. Provided are the percent identities, percent similarity, and the percent gap. The alignment of the sequences is provided in Figure 22.



#### 4C. Phytyl Prenyltransferase Enzyme Assays

[<sup>3</sup>H] Homogentisic acid in 0.1% H<sub>3</sub>PO<sub>4</sub> (specific radioactivity 40 Ci/mmol).

Phytyl pyrophosphate was synthesized as described by Joo, *et al.* (1973) *Can J.*

5 *Biochem.* 51:1527. 2-methyl-6-phytylquinol and 2,3-dimethyl-5-phytylquinol were synthesized as described by Soll, *et al.* (1980) *Phytochemistry* 19:215. Homogentisic acid,  $\alpha$ ,  $\beta$ ,  $\delta$ , and  $\gamma$ -tocopherol, and tocol, were purchased commercially.

The wild-type strain of *Synechocystis* sp. PCC 6803 was grown in BG11 medium with bubbling air at 30°C under 50  $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$  fluorescent light, and 70% relative  
10 humidity. The growth medium of slr1736 knock-out (potential PPT) strain of this organism was supplemented with 25  $\mu\text{g mL}^{-1}$  kanamycin. Cells were collected from 0.25 to 1 liter culture by centrifugation at 5000 *g* for 10 min and stored at -80°C.

Total membranes were isolated according to Zak's procedures with some modifications (Zak, *et al.* (1999) *Eur J. Biochem* 261:311). Cells were broken on a  
15 French press. Before the French press treatment, the cells were incubated for 1 hour with lysozyme (0.5%, w/v) at 30 °C in a medium containing 7 mM EDTA, 5 mM NaCl and 10 mM Hepes-NaOH, pH 7.4. The spheroplasts were collected by centrifugation at 5000 *g* for 10 min and resuspended at 0.1 - 0.5 mg chlorophyll·mL<sup>-1</sup> in 20 mM potassium phosphate buffer, pH 7.8. Proper amount of protease inhibitor cocktail and  
20 DNAase I from Boehringer Mannheim were added to the solution. French press treatments were performed two to three times at 100 MPa. After breakage, the cell suspension was centrifuged for 10 min at 5000*g* to pellet unbroken cells, and this was followed by centrifugation at 100 000 *g* for 1 hour to collect total membranes. The final pellet was resuspended in a buffer containing 50 mM Tris-HCL and 4 mM MgCl<sub>2</sub>.

25 Chloroplast pellets were isolated from 250 g of spinach leaves obtained from local markets. Devined leaf sections were cut into grinding buffer (2 l /250 g leaves) containing 2 mM EDTA, 1 mM MgCl<sub>2</sub>, 1 mM MnCl<sub>2</sub>, 0.33 M sorbitol, 0.1% ascorbic acid, and 50 mM Hepes at pH 7.5. The leaves were homogenized for 3 sec three times in a 1-L blender, and filtered through 4 layers of miracloth. The supernatant was then

centrifuged at 5000g for 6 min. The chloroplast pellets were resuspended in small amount of grinding buffer (Douce, *et al* Methods in Chloroplast Molecular Biology, 239 (1982)

Chloroplasts in pellets can be broken in three ways. Chloroplast pellets were first  
5 aliquoted in 1 mg of chlorophyll per tube, centrifuged at 6000 rpm for 2 min in microcentrifuge, and grinding buffer was removed. Two hundred microliters of Triton X-100 buffer (0.1% Triton X-100, 50 mM Tris-HCl pH 7.6 and 4 mM MgCl<sub>2</sub>) or swelling buffer (10 mM Tris pH 7.6 and 4 mM MgCl<sub>2</sub>) was added to each tube and incubated for ½ hour at 4°C. Then the broken chloroplast pellets were used for the  
10 assay immediately. In addition, broken chloroplasts can also be obtained by freezing in liquid nitrogen and stored at -80°C for ½ hour, then used for the assay.

In some cases chloroplast pellets were further purified with 40%/ 80% percoll gradient to obtain intact chloroplasts. The intact chloroplasts were broken with swelling buffer, then either used for assay or further purified for envelope membranes  
15 with 20.5%/ 31.8% sucrose density gradient (Sol, *et al* (1980) *supra*). The membrane fractions were centrifuged at 100 000g for 40 min and resuspended in 50 mM Tris-HCl pH 7.6, 4 mM MgCl<sub>2</sub>.

Various amounts of [<sup>3</sup>H]HGA, 40 to 60 µM unlabelled HGA with specific activity in the range of 0.16 to 4 Ci/mmol were mixed with a proper amount of 1M Tris-  
20 NaOH pH 10 to adjust pH to 7.6. HGA was reduced for 4 min with a trace amount of solid NaBH<sub>4</sub>. In addition to HGA, standard incubation mixture (final vol 1 mL) contained 50 mM Tris-HCl, pH 7.6, 3-5 mM MgCl<sub>2</sub>, and 100 µM phytyl pyrophosphate. The reaction was initiated by addition of *Synechocystis* total membranes, spinach chloroplast pellets, spinach broken chloroplasts, or spinach  
25 envelope membranes. The enzyme reaction was carried out for 2 hour at 23°C or 30°C in the dark or light. The reaction is stopped by freezing with liquid nitrogen, and stored at -80°C or directly by extraction.

A constant amount of tocol was added to each assay mixture and reaction products were extracted with a 2 mL mixture of chloroform/methanol (1:2, v/v) to give a



monophasic solution. NaCl solution (2 mL; 0.9%) was added with vigorous shaking. This extraction procedure was repeated three times. The organic layer containing the prenylquinones was filtered through a 20 mμ filter, evaporated under N<sub>2</sub> and then resuspended in 100 μL of ethanol.

- 5        The samples were mainly analyzed by Normal-Phase HPLC method (Isocratic 90% Hexane and 10% Methyl-t-butyl ether), and use a Zorbax silica column, 4.6 x 250 mm. The samples were also analyzed by Reversed-Phase HPLC method (Isocratic 0.1% H<sub>3</sub>PO<sub>4</sub> in MeOH), and use a Vydac 201HS54 C18 column; 4.6 x 250 mm coupled with an All-tech C18 guard column. The amount of products were calculated  
10    based on the substrate specific radioactivity, and adjusted according to the % recovery based on the amount of internal standard.

The amount of chlorophyll was determined as described in Arnon (1949) *Plant Physiol.* 24:1. Amount of protein was determined by the Bradford method using gamma globulin as a standard (Bradford, (1976) *Anal. Biochem.* 72:248)

- 15        Results of the assay demonstrate that 2-Methyl-6-Phytylplastoquinone is not produced in the *Synechocystis* slr1736 knockout preparations. The results of the phytyl prenyltransferase enzyme activity assay for the slr1736 knock out are presented in Figure 23.

20    4D. Complementation of the slr1736 knockout with ATPT2

- In order to determine whether ATPT2 could complement the knockout of slr1736 in *Synechocystis* 6803, a plasmid was constructed to express the ATPT2 sequence from the TAC promoter. A vector, plasmid psl1211, was obtained from the lab of Dr. Himadri Pakrasi of Washington University, and is based on the plasmid  
25    RSF1010 which is a broad host range plasmid (Ng W.-O., Zentella R., Wang, Y., Taylor J-S. A., Pakrasi, H.B. 2000. *phrA*, the major photoreactivating factor in the cyanobacterium *Synechocystis* sp. strain PCC 6803 codes for a cyclobutane pyrimidine dimer specific DNA photolyase. *Arch. Microbiol.* (in press)). The ATPT2 gene was isolated from the vector pCGN10817 by PCR using the following primers.

ATPT2nco.pr 5'-CCATGGATTCGAGTAAAGTTGTCGC (SEQ ID NO:89);  
ATPT2ri.pr- 5'-GAATTCACCTTCAAAAAAGGTAACAG (SEQ ID NO:90). These  
primers will remove approximately 112 BP from the 5' end of the ATPT2 sequence,  
which is thought to be the chloroplast transit peptide. These primers will also add an  
5 NcoI site at the 5' end and an EcoRI site at the 3' end which can be used for sub-  
cloning into subsequent vectors. The PCR product from using these primers and  
pCGN10817 was ligated into pGEM T easy and the resulting vector pMON21689 was  
confirmed by sequencing using the m13forward and m13reverse primers. The  
NcoI/EcoRI fragment from pMON21689 was then ligated with the EagI/EcoRI and  
10 EagI/NcoI fragments from psl1211 resulting in pMON21690. The plasmid  
pMON21690 was introduced into the slr1736 *Synechocystis* 6803 KO strain via  
conjugation. Cells of sl906 (a helper strain) and DH10B cells containing  
pMON21690 were grown to log phase (O.D. 600= 0.4) and 1 ml was harvested by  
centrifugation. The cell pellets were washed twice with a sterile BG-11 solution and  
15 resuspended in 200 ul of BG-11. The following was mixed in a sterile eppendorf  
tube: 50 ul SL906, 50 ul DH10B cells containing pMON21690, and 100 ul of a fresh  
culture of the slr1736 *Synechocystis* 6803 KO strain (O.D. 730 = 0.2-0.4). The cell  
mixture was immediately transferred to a nitrocellulose filter resting on BG-11 and  
incubated for 24 hours at 30C and 2500 LUX(50 ue) of light. The filter was then  
20 transferred to BG-11 supplemented with 10ug/ml Gentamycin and incubated as above  
for ~5 days. When colonies appeared, they were picked and grown up in liquid BG-  
11 + Gentamycin 10 ug/ml. (Elhai, J. and Wolk, P. 1988. Conjugal transfer of DNA  
to Cyanobacteria. *Methods in Enzymology* 167, 747-54) The liquid cultures were then  
assayed for tocopherols by harvesting 1ml of culture by centrifugation, extracting with  
25 ethanol/pyrogallol, and HPLC separation. The slr1736 *Synechocystis* 6803 KO strain,  
did not contain any detectable tocopherols, while the slr1736 *Synechocystis* 6803 KO  
strain transformed with pmon21690 contained detectable alpha tocopherol. A  
*Synechocystis* 6803 strain transformed with psl1211(vector control) produced alpha  
tocopherol as well.

#### 4E: Additional Evidence of Prenyltransferase Activity

To test the hypothesis that slr1736 or ATPT2 are sufficient as single genes to obtain phytyl prenyltransferase activity, both genes were expressed in SF9 cells and in yeast. When either slr1736 or ATPT2 were expressed in insect cells (Table 5) or in yeast, phytyl prenyltransferase activity was detectable in membrane preparations, whereas membrane preparations of the yeast vector control, or membrane preparations of insect cells did not exhibit phytyl prenyltransferase activity.

Table 5: Phytyl prenyltransferase activity

Enzyme source	Enzyme activity [pmol/mg x h]
slr1736 expressed in SF9 cells	20
ATPT2 expressed in SF9 cells	6
SF9 cell control	< 0.05
<i>Synechocystis</i> 6803	0.25
Spinach chloroplasts	0.20

#### Example 5: Transgenic Plant Analysis

##### 5A. *Arabidopsis*

*Arabidopsis* plants transformed with constructs for the sense or antisense expression of the ATPT proteins were analyzed by High Pressure Liquid Chromatography (HPLC) for altered levels of total tocopherols, as well as altered levels of specific tocopherols (alpha, beta, gamma, and delta tocopherol).

Extracts of leaves and seeds were prepared for HPLC as follows. For seed extracts, 10 mg of seed was added to 1 g of microbeads (Biospec) in a sterile microfuge tube to which 500 ul 1% pyrogallol (Sigma Chem)/ethanol was added. The mixture was shaken for 3 minutes in a mini Beadbeater (Biospec) on "fast" speed.

The extract was filtered through a 0.2  $\mu$ m filter into an autosampler tube. The filtered extracts were then used in HPLC analysis described below.

Leaf extracts were prepared by mixing 30-50 mg of leaf tissue with 1 g microbeads and freezing in liquid nitrogen until extraction. For extraction, 500  $\mu$ l 1% pyrogallol in ethanol was added to the leaf/bead mixture and shaken for 1 minute on a Beadbeater (Biospec) on "fast" speed. The resulting mixture was centrifuged for 4 minutes at 14,000 rpm and filtered as described above prior to HPLC analysis.

HPLC was performed on a Zorbax silica HPLC column (4.6 mm X 250 mm) with a fluorescent detection, an excitation at 290 nm, an emission at 336 nm, and bandpass and slits. Solvent A was hexane and solvent B was methyl-t-butyl ether. The injection volume was 20  $\mu$ l, the flow rate was 1.5 ml/min, the run time was 12 min (40°C) using the gradient (Table 6):

**Table 6:**

<u>Time</u>	<u>Solvent A</u>	<u>Solvent B</u>
0 min.	90%	10%
10 min.	90%	10%
11 min.	25%	75%
12 min.	90%	10%

Tocopherol standards in 1% pyrogallol/ ethanol were also run for comparison (alpha tocopherol, gamma tocopherol, beta tocopherol, delta tocopherol, and tocopherol (tocol) (all from Matreya).

Standard curves for alpha, beta, delta, and gamma tocopherol were calculated using Chemstation software. The absolute amount of component x is: Absolute amount of x =  $\text{Response}_x \times \text{RF}_x \times \text{dilution factor}$  where  $\text{Response}_x$  is the area of peak x,  $\text{RF}_x$  is the response factor for component x ( $\text{Amount}_x / \text{Response}_x$ ) and the dilution factor is 500  $\mu$ l. The ng/mg tissue is found by: total ng component/mg plant tissue.

Results of the HPLC analysis of seed extracts of transgenic *Arabidopsis* lines containing pMON10822 for the expression of ATPT2 from the napin promoter are provided in Figure 24.

HPLC analysis results of segregating T2 *Arabidopsis* seed tissue expressing  
5 the ATPT2 sequence from the napin promoter (pCGN10822) demonstrates an increased level of tocopherols in the seed. Total tocopherol levels are increased as much as 50% over the total tocopherol levels of non-transformed (wild-type) *Arabidopsis* plants (Figure 25). Homozygous progeny from the top 3 lines (T3 seed) have up to a two-fold (100%) increase in total tocopherol levels over control  
10 *Arabidopsis* seed ( Figure 26.)

Furthermore, increases of particular tocopherols are also increased in transgenic *Arabidopsis* plants expressing the ATPT2 nucleic acid sequence from the napin promoter. Levels of delta tocopherol in these lines are increased greater than 3 fold over the delta tocopherol levels obtained from the seeds of wild type *Arabidopsis*  
15 lines. Levels of gamma tocopherol in transgenic *Arabidopsis* lines expressing the ATPT2 nucleic acid sequence are increased as much as about 60% over the levels obtained in the seeds of non-transgenic control lines. Furthermore, levels of alpha tocopherol are increased as much as 3 fold over those obtained from non-transgenic control lines.

20 Results of the HPLC analysis of seed extracts of transgenic *Arabidopsis* lines containing pCGN10803 for the expression of ATPT2 from the enhanced 35S promoter (antisense orientation ) are provided in Figure 25. Two lines were identified that have reduced total tocopherols, up to a ten-fold decrease observed in T3 seed compared to control *Arabidopsis* (Figure 27.)

25

#### 5B. Canola

Brassica napus, variety SP30021, was transformed with pCGN10822 (napin-ATPT2-napin 3', sense orientation) using *Agrobacterium tumefaciens*-mediated

transformation. Flowers of the R0 plants were tagged upon pollination and developing seed was collected at 35 and 45 days after pollination (DAP).

Developing seed was assayed for tocopherol levels, as described above for *Arabidopsis*. Line 10822-1 shows a 20% increase of total tocopherols, compared to the wild-type control, at 45 DAP. Figure 28 shows total tocopherol levels measured in developing canola seed.

#### **Example 6: Sequences to Tocopherol Cyclase**

##### **6A. Preparation of the slr1737 Knockout**

10       The *Synechocystis* sp. 6803 slr1737 knockout was constructed by the following method. The GPS™-1 Genome Priming System (New England Biolabs) was used to insert, by a Tn7 Transposase system, a Kanamycin resistance cassette into *slr1737*. A plasmid from a *Synechocystis* genomic library clone containing 652 base pairs of the targeted orf (*Synechocystis* genome base pairs 1324051 – 1324703; 15 the predicted orf base pairs 1323672 – 1324763, as annotated by Cyanobase) was used as target DNA. The reaction was performed according to the manufacturers protocol. The reaction mixture was then transformed into *E. coli* DH10B electrocompetant cells and plated. Colonies from this transformation were then screened for transposon insertions into the target sequence by amplifying with M13 Forward and Reverse 20 Universal primers, yielding a product of 652 base pairs plus ~1700 base pairs, the size of the transposon kanamycin cassette, for a total fragment size of ~2300 base pairs. After this determination, it was then necessary to determine the approximate location of the insertion within the targeted orf, as 100 base pairs of orf sequence was estimated as necessary for efficient homologous recombination in *Synechocystis*. This 25 was accomplished through amplification reactions using either of the primers to the ends of the transposon, Primer S (5' end) or N (3' end), in combination with either a M13 Forward or Reverse primer. That is, four different primer combinations were used to map each potential knockout construct: Primer S – M13 Forward, Primer S – M13 Reverse, Primer N – M13 Forward, Primer N – M13 Reverse. The construct

used to transform *Synechocystis* and knockout *slr1737* was determined to consist of a approximately 150 base pairs of *slr1737* sequence on the 5' side of the transposon insertion and approximately 500 base pairs on the 3' side, with the transcription of the orf and kanamycin cassette in the same direction. The nucleic acid sequence of  
5 *slr1737* is provided in SEQ ID NO:38 the deduced amino acid sequence is provided in SEQ ID NO:39.

Cells of *Synechocystis* 6803 were grown to a density of  $\sim 2 \times 10^8$  cells per ml and harvested by centrifugation. The cell pellet was re-suspended in fresh BG-11 medium at a density of  $1 \times 10^9$  cells per ml and used immediately for transformation.  
10 100 ul of these cells were mixed with 5 ul of mini prep DNA and incubated with light at 30C for 4 hours. This mixture was then plated onto nylon filters resting on BG-11 agar supplemented with TES pH8 and allowed to grow for 12-18 hours. The filters were then transferred to BG-11 agar + TES + 5ug/ml kanamycin and allowed to grow until colonies appeared within 7-10 days (Packer and Glazer, 1988). Colonies were  
15 then picked into BG-11 liquid media containing 5 ug/ml kanamycin and allowed to grow for 5 days. These cells were then transferred to Bg-11 media containing 10ug/ml kanamycin and allowed to grow for 5 days and then transferred to Bg-11 + kanamycin at 25ug/ml and allowed to grow for 5 days. Cells were then harvested for PCR analysis to determine the presence of a disrupted ORF and also for HPLC  
20 analysis to determine if the disruption had any effect on tocopherol levels.

PCR analysis of the *Synechocystis* isolates, using primers to the ends of the *slr1737* orf, showed complete segregation of the mutant genome, meaning no copies of the wild type genome could be detected in these strains. This suggests that function of the native gene is not essential for cell function. HPLC analysis of the  
25 strain carrying the knockout for *slr1737* produced no detectable levels of tocopherol.

#### 6B. The relation of *slr1737* and *slr1736*

The *slr1737* gene occurs in *Synechocystis* downstream and in the same orientation as *slr1736*, the phytyl prenyltransferase. In bacteria this proximity often



indicates an operon structure and therefore an expression pattern that is linked in all genes belonging to this operon. Occasionally such operons contain several genes that are required to constitute one enzyme. To confirm that slr1737 is not required for phytyl prenyltransferase activity, phytyl prenyltransferase was measured in extracts  
5 from the *Synechocystis* slr1737 knockout mutant. Figure 29 shows that extracts from the *Synechocystis* slr1737 knockout mutant still contain phytyl prenyltransferase activity. The molecular organization of genes in *Synechocystis* 6803 is shown in A. Figures B and C show HPLC traces (normal phase HPLC) of reaction products obtained with membrane preparations from *Synechocystis* wild type and slr1737  
10 membrane preparations, respectively.

The fact that slr1737 is not required for the PPT activity provides additional data that ATPT2 and slr1736 encode phytyl prenyltransferases.

#### 6C *Synechocystis* Knockouts

15 *Synechocystis* 6803 wild type and *Synechocystis* slr1737 knockout mutant were grown photoautotrophically. Cells from a 20 ml culture of the late logarithmic growth phase were harvested and extracted with ethanol. Extracts were separated by isocratic normal-phase HPLC using a Hexane/Methyl-t-butyl ether (95/5) and a Zorbax silica column, 4.6 x 250 mm. Tocopherols and tocopherol intermediates were  
20 detected by fluorescence (excitement 290 nm, emission 336 nm) (Figure 30).

Extracts of *Synechocystis* 6803 contained a clear signal of alpha-tocopherol. 2,3-Dimethyl-5-phytylplastoquinol was below the limit of detection in extracts from the *Synechocystis* wild type (C). In contrast, extracts from the *Synechocystis* slr1737 knockout mutant did not contain alpha-tocopherol, but contained 2,3-dimethyl-5-  
25 phytylplastoquinol (D), indicating that the interruption of slr1737 has resulted in a block of the 2,3-dimethyl-5-phytylplastoquinol cyclase reaction.

Chromatograms of standard compounds alpha, beta, gamma, delta-tocopherol and 2,3-dimethyl-5-phytylplastoquinol are shown in A and B. Chromatograms of extracts from *Synechocystis* wild type and the *Synechocystis* slr1737 knockout mutant

are shown in C and D, respectively. Abbreviations: 2,3-DMPQ, 2,3-dimethyl-5-phytylplastoquinol.

#### 6D. Incubation with Lysozyme treated *Synechocystis*

5        *Synechocystis* 6803 wild type and slr1737 knockout mutant cells from the late logarithmic growth phase (approximately 1 g wet cells per experiment in a total volume of 3 ml) were treated with Lysozyme and subsequently incubated with S-adenosylmethionine, and phytylpyrophosphate, plus radiolabelled homogentisic acid. After 17h incubation in the dark at room temperature the samples were extracted  
10 with 6 ml chloroform / methanol (1/2 v/v). Phase separation was obtained by the addition of 6 ml 0.9% NaCl solution. This procedure was repeated three times. Under these conditions 2,3-dimethyl-5-phytylplastoquinol is oxidized to form 2,3-dimethyl-5-phytylplastoquinone.

The extracts were analyzed by normal phase and reverse phase HPLC. Using  
15 extracts from wild type *Synechocystis* cells radiolabelled gamma-tocopherol and traces of radiolabelled 2,3-dimethyl-5-phytylplastoquinone were detected. When extracts from the slr1737 knockout mutant were analyzed, only radiolabelled 2,3-dimethyl-5-phytylplastoquinone was detectable. The amount of 2,3-dimethyl-5-phytylplastoquinone was significantly increased compared to wild type extracts. Heat  
20 treated samples of the wild type and the slr1737 knockout mutant did not produce radiolabelled 2,3-dimethyl-5-phytylplastoquinone, nor radiolabelled tocopherols. These results further support the role of the slr1737 expression product in the cyclization of 2,3-dimethyl-5-phytylplastoquinol.

#### 25 6E. *Arabidopsis* Homologue to slr1737

An *Arabidopsis* homologue to slr1737 was identified from a BLASTALL search using *Synechocystis* sp 6803 gene slr1737 as the query, in both public and proprietary databases. SEQ ID NO:109 and SEQ ID NO:110 are the DNA and

translated amino acid sequences, respectively, of the *Arabidopsis* homologue to slr1737. The start is found at the ATG at base 56 in SEQ ID NO:109.

The sequences obtained for the homologue from the proprietary database differs from the public database (F4D11.30, BAC AL022537), in having a start site  
5 471 base pairs upstream of the start identified in the public sequence. A comparison of the public and proprietary sequences is provided in Figure 31. The correct start correlates within the public database sequence is at 12080, while the public sequence start is given as being at 11609.

Attempts to amplify a slr1737 homologue were unsuccessful using primers  
10 designed from the public database, while amplification of the gene was accomplished with primers obtained from SEQ ID NO:109.

Analysis of the protein sequence to identify transit peptide sequence predicted two potential cleavage sites, one between amino acids 48 and 49, and the other between amino acids 98 and 99.

15

#### 6F. slr1737 Protein Information

The slr1737 orf comprises 363 amino acid residues and has a predicted MW of 41kDa (SEQ ID NO: 39). Hydropathic analysis indicates the protein is hydrophilic (Figure 32).

20

The *Arabidopsis* homologue to slr1737 (SEQ ID xx) comprises 488 amino acid residues, has a predicted MW of 55kDa, and has a putative transit peptide sequence comprising the first 98 amino acids. The predicted MW of the mature form of the *Arabidopsis* homologue is 44kDa. The hydropathic plot for the *Arabidopsis* homologue also reveals that it is hydrophilic (Figure 33). Further blast analysis of  
25 the *Arabidopsis* homologue reveals limited sequence identity (25 % sequence identity) with the beta-subunit of respiratory nitrate reductase. Based on the sequence identity to nitrate reductase, it suggests the slr1737 orf is an enzyme that likely involves general acid catalysis mechanism.

Investigation of known enzymes involved in tocopherol metabolism indicated that the best candidate corresponding to the general acid mechanism is the tocopherol cyclase. There are many known examples of cyclases including, tocopherol cyclase, chalcone isomerase, lycopene cyclase, and aristolochene synthase. By further  
5 examination of the microscopic catalytic mechanism of phytoplastoquinol cyclization, as an example, chalcone isomerase has a catalytic mechanism most similar to tocopherol cyclase. (Figure 34).

Multiple sequence alignment was performed between slr1737, slr1737  
*Arabidopsis* homologue and the *Arabidopsis* chalcone isomerase (Genbank:P41088)  
10 (Figure 35). 65% of the conserved residues among the three enzymes are strictly conserved within the known chalcone isomerases. The crystal structure of alfalfa chalcone isomerase has been solved (Jez, Joseph M., Bowman, Marianne E., Dixon, Richard A., and Noel, Joseph P. (2000) "Structure and mechanism of the  
evolutionarily unique plant enzyme chalcone isomerase". *Nature Structural Biology*  
15 7: 786-791.) It has been demonstrated tyrosine (Y) 106 of the alfalfa chalcone isomerase serves as the general acid during cyclization reaction (Genbank: P28012). The equivalent residue in slr1737 and the slr1737 *Arabidopsis* homolog is lysine (K), which is an excellent catalytic residue as general acid.

The information available from partial purification of tocopherol cyclase from  
20 *Chlorella protothecoides* (U.S. Patent No. 5,432,069), *i.e.*, described as being glycine rich, water soluble and with a predicted MW of 48-50kDa, is consistent with the protein informatics information obtained for the slr1737 and the *Arabidopsis* slr1737 homologue.

All publications and patent applications mentioned in this specification are  
25 indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claim.

## CLAIMS

What is claimed is:

1. An isolated nucleic acid sequence encoding a tocopherol cyclase.
- 5 2. An isolated nucleic acid sequence according to Claim 1, wherein said tocopherol cyclase is active in the cyclization of 2,3-dimethyl-5-phytylplastoquinol to tocopherol.
3. An isolated nucleic acid sequence according to Claim 1, wherein said tocopherol cyclase is active in the cyclization of 2,3-dimethyl-5-geranylgeranylplastoquinol to tocotrienol.
- 10 4. An isolated DNA sequence according to Claim 1, wherein said nucleic acid sequence is isolated from a eukaryotic cell source.
5. An isolated DNA sequence according to Claim 4, wherein said eukaryotic cell source is selected from the group consisting of mammalian, nematode, fungal, and plant cells.
6. The DNA encoding sequence of Claim 5 wherein said tocopherol cyclase protein is from  
15 *Arabidopsis*.
7. The DNA encoding sequence of Claim 6 wherein said tocopherol cyclase protein is encoded by a sequence of SEQ ID NO:109.
8. The DNA encoding sequence of Claim 7 wherein said tocopherol cyclase protein has an amino acid sequence of SEQ ID NO:110.
- 20 9. The DNA encoding sequence of Claim 4 wherein said tocopherol cyclase protein is from a source selected from the group consisting of *Arabidopsis*, soybean, corn, rice, wheat, leek canola, , leek, cotton, and tomato.
10. An isolated DNA sequence according to Claim 4, wherein said prokaryotic source is a *Synechocystis* sp.
- 25 11. The DNA encoding sequence of Claim 10 wherein said tocopherol cyclase protein is encoded by a sequence of SEQ ID NO:38.
12. The DNA encoding sequence of Claim 10 wherein said tocopherol cyclase protein has an amino acid sequence of SEQ ID NO:39.

13. A nucleic acid construct comprising as operably linked components, a transcriptional initiation region functional in a host cell, a nucleic acid sequence encoding a tocopherol cyclase, and a transcriptional termination region.
14. A nucleic acid construct according to Claim 13, wherein said nucleic acid sequence  
5 encoding tocopherol cyclase is obtained from an organism selected from the group consisting of a eukaryotic organism and a prokaryotic organism.
15. A nucleic acid construct according to Claim 14, wherein said nucleic acid sequence encoding tocopherol cyclase is obtained from a plant source.
16. A nucleic acid construct according to Claim 15, wherein said nucleic acid sequence  
10 encoding tocopherol cyclase is obtained from a source selected from the group consisting of *Arabidopsis*, soybean, corn, rice, wheat, leek canola, , leek, cotton, and tomato.
17. A nucleic acid construct according to Claim 13, wherein said nucleic acid sequence encoding tocopherol cyclase is obtained from a *Synechocystis* sp.
18. A plant cell comprising the construct of 13.
- 15 19. A plant comprising a cell of Claim 18.
- 20 A feed composition produced from a plant according to Claim 19.
21. A seed comprising a cell of Claim 18.
- 22 Oil obtained from a seed of Claim 21.
23. A natural tocopherol rich refined and deodorised oil which has been produced by  
20 a method of treating an oil according to Claim 22 by distilling under low pressure and high temperature, wherein said refined oil has reduced free fatty acids and a substantial percentage of tocopherol present in the pretreated oil.
24. A refined oil according to claim 23, wherein the pretreated oil is crude or pre-treated soybean oil.
- 25 25. A refined oil according to claim 23, wherein the refined oil is degummed and bleached.
26. A method for the alteration of the isoprenoid content in a host cell, said method comprising; transforming said host cell with a construct comprising as operably linked



components, a transcriptional initiation region functional in a host cell, a nucleic acid sequence encoding tocopherol cyclase, and a transcriptional termination region, wherein said isoprenoid compound selected from the group of tocopherols and tocotrienols.

27. The method according to Claim 26, wherein said host cell is selected from the group consisting of a prokaryotic cell and a eukaryotic cell.

28. The method according to Claim 27, wherein said prokaryotic cell is a *Synechocystis* sp.

29. The method according to Claim 27, wherein said eukaryotic cell is a plant cell.

30. The method according to Claim 29, wherein said plant cell is obtained from a plant selected from the group consisting of *Arabidopsis*, soybean, corn, rice, wheat, leek canola, , leek, cotton, and tomato.

31. A method for producing an isoprenoid compound of interest in a host cell, said method comprising obtaining a transformed host cell, said host cell having and expressing in its genome:

a construct having a DNA sequence encoding a tocopherol cyclase operably linked to a transcriptional initiation region functional in a host cell, wherein said isoprenoid compound selected from the group of tocopherols and tocotrienols.

32. The method according to Claim 31, wherein said host cell is selected from the group consisting of a prokaryotic cell and a eukaryotic cell.

33. The method according to Claim 32, wherein said prokaryotic cell is a *Synechocystis* sp.

34. The method according to Claim 32, wherein said eukaryotic cell is a plant cell.

35. The method according to Claim 34, wherein said plant cell is obtained from a plant selected from the group consisting wherein said compound selected from the group of *Arabidopsis*, soybean, corn, rice, wheat, leek canola, , leek, cotton, and tomato.

36. A method for increasing the biosynthetic flux in a host cell toward production of an isoprenoid compound, said method comprising; transforming said host cell with a construct comprising as operably linked components, a transcriptional initiation region functional in a host cell, a DNA encoding a tocopherol cyclase, and a transcriptional termination region, wherein said isoprenoid compound selected from the group of tocopherols and tocotrienols,.

37. The method according to Claim 36, wherein said host cell is selected from the group consisting of a prokaryotic cell and a eukaryotic cell.
38. The method according to Claim 37, wherein said prokaryotic cell is a *Synechocystis* sp.
39. The method according to Claim 37, wherein said eukaryotic cell is a plant cell.
- 5 40. The method according to Claim 39, wherein said plant cell is obtained from a plant selected from the group consisting *Arabidopsis*, soybean, corn, rice, wheat, leek canola, , leek, cotton, and tomato.
41. The method according to Claim 39, wherein said transcriptional initiation region is a seed-specific promoter.

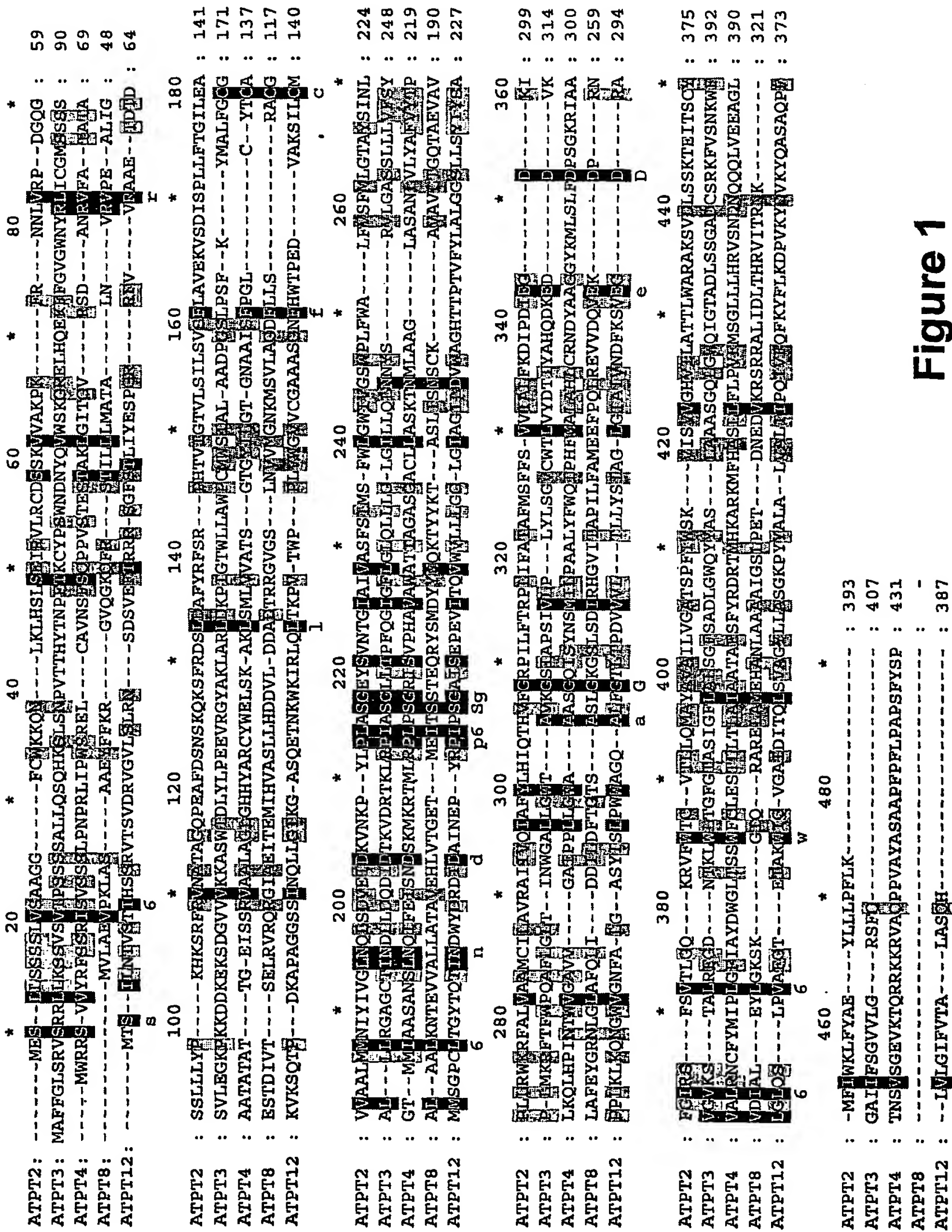


Figure 1

2/40

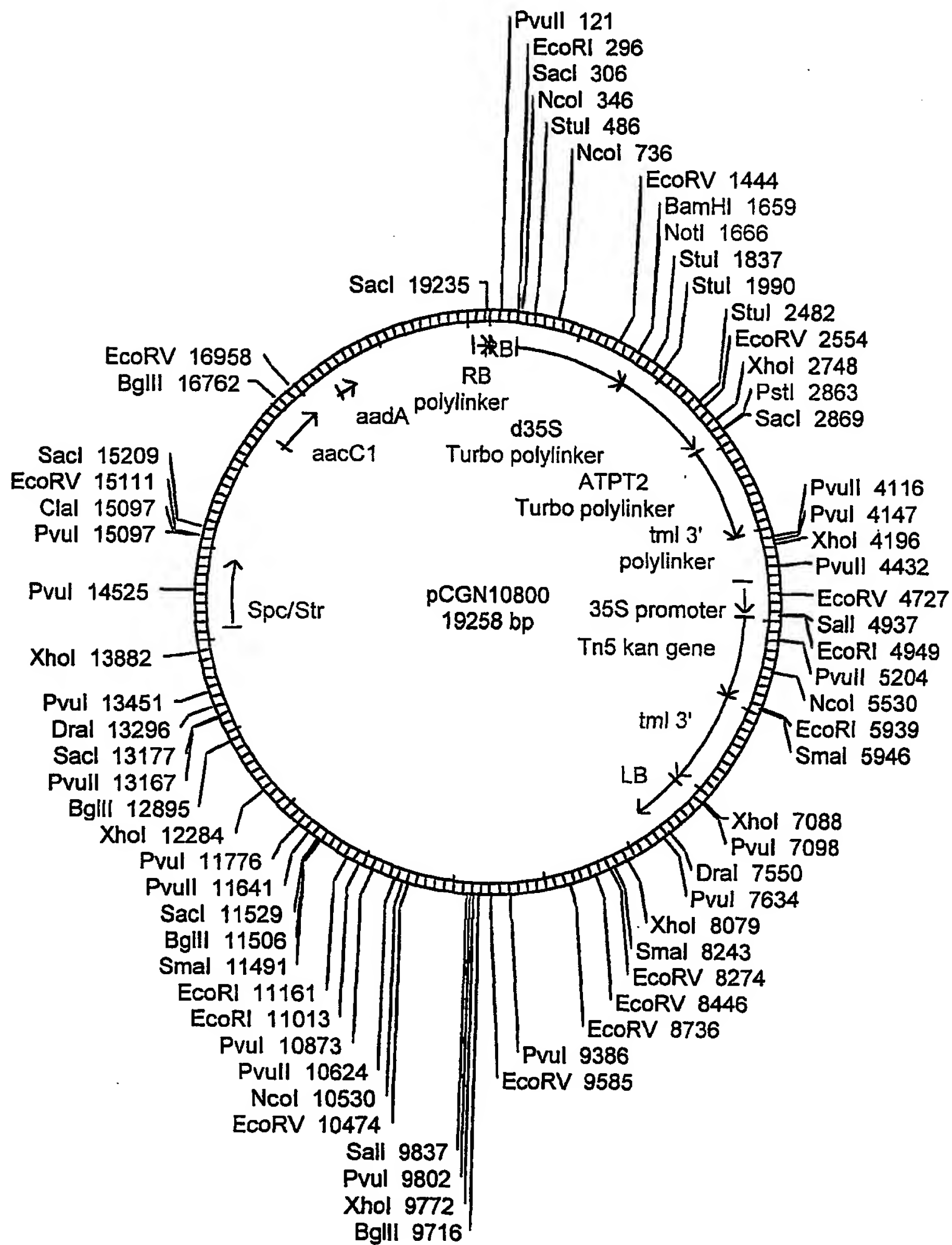


Figure 2

3/40

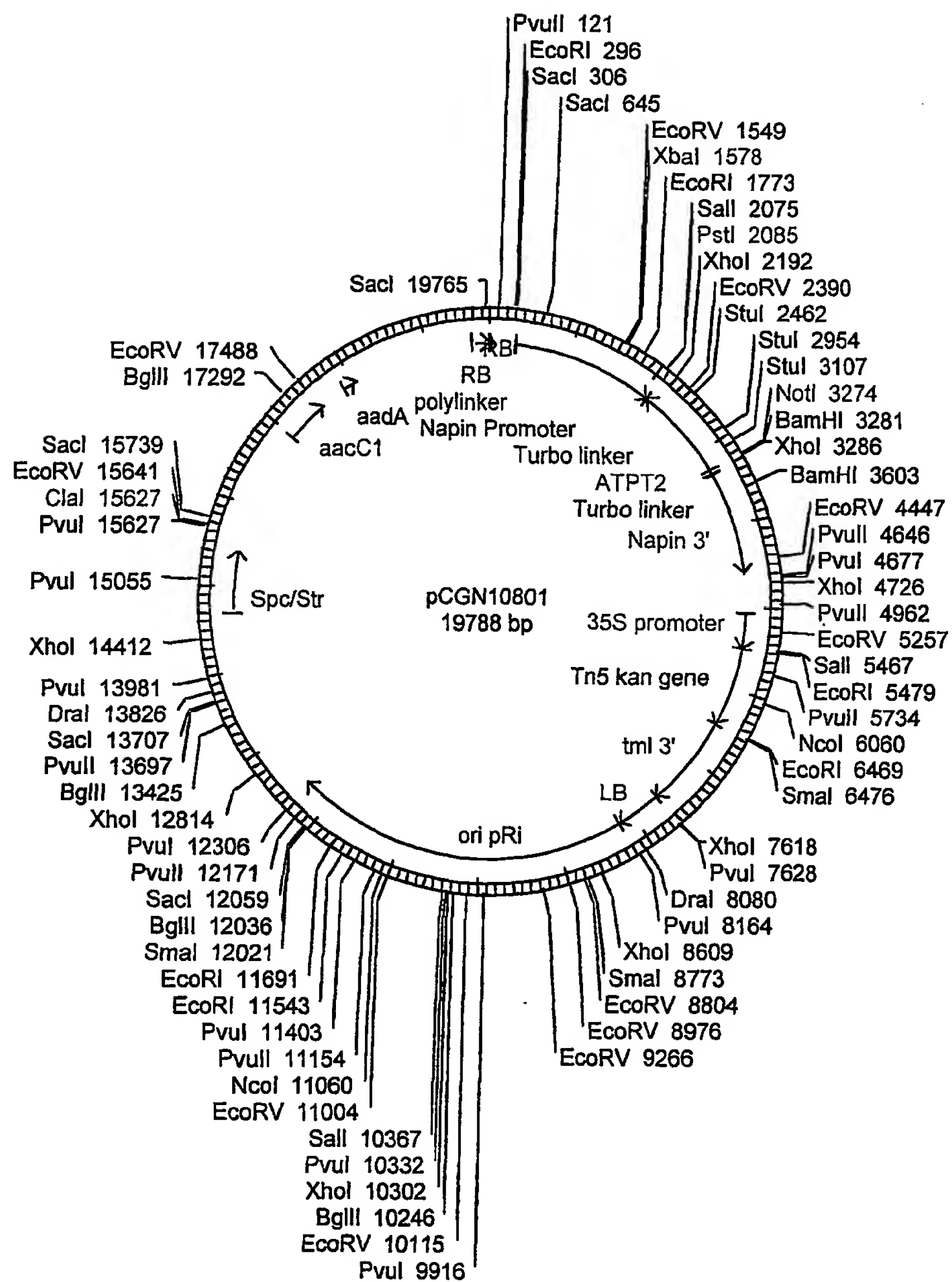


Figure 3

4/40

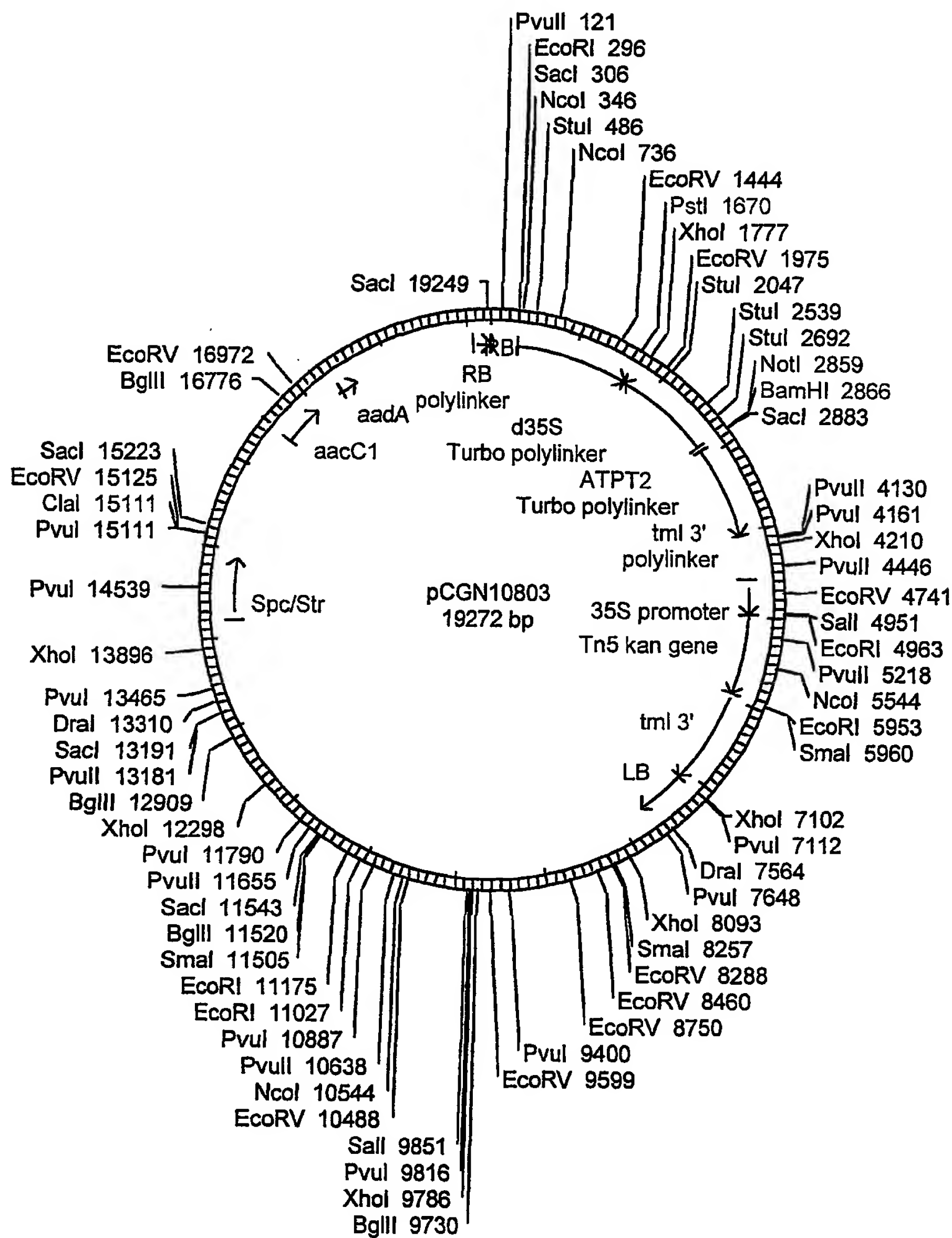


Figure 4

5/40

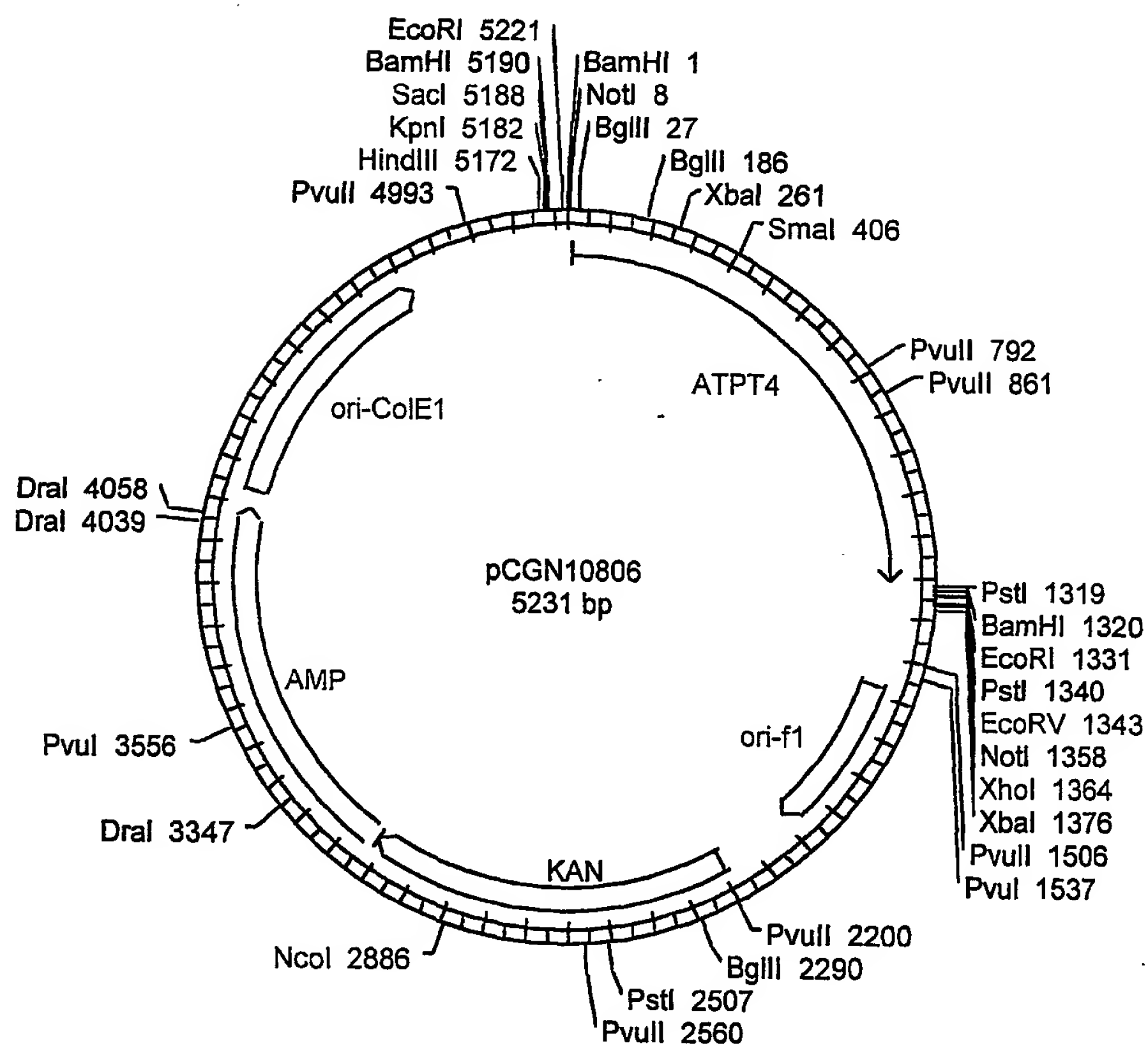


Figure 5



6/40

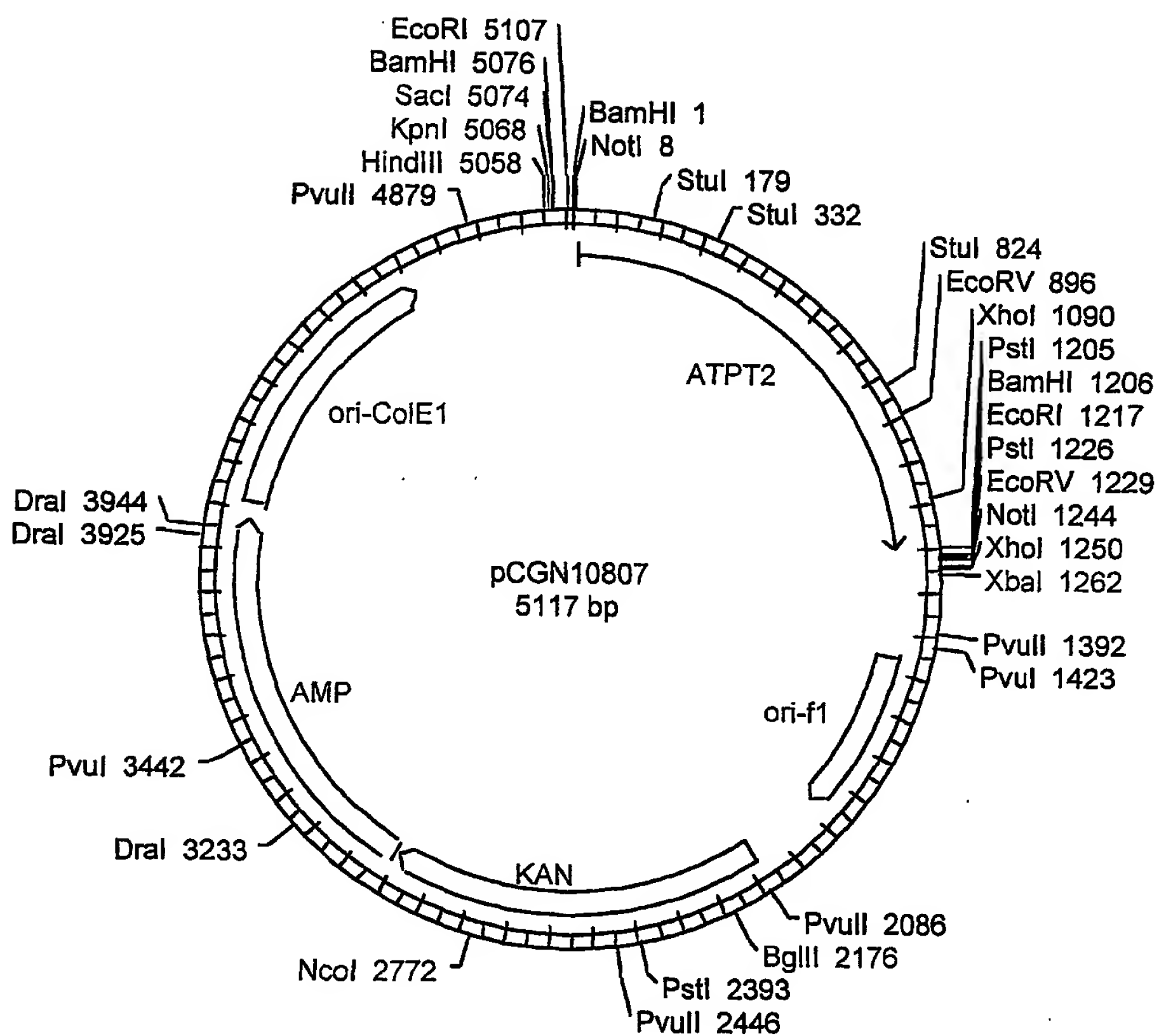


Figure 6

7/40

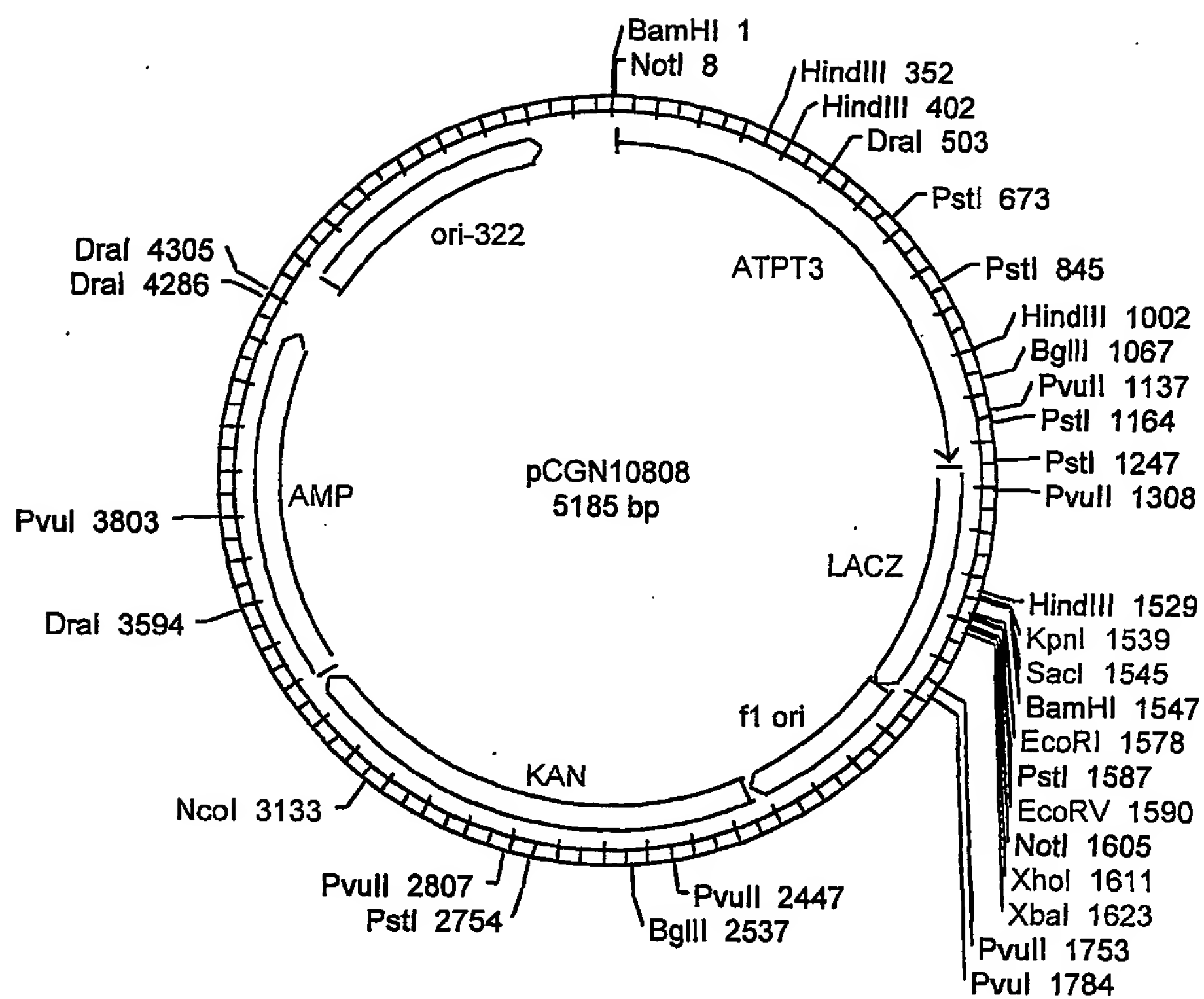


Figure 7

8/40

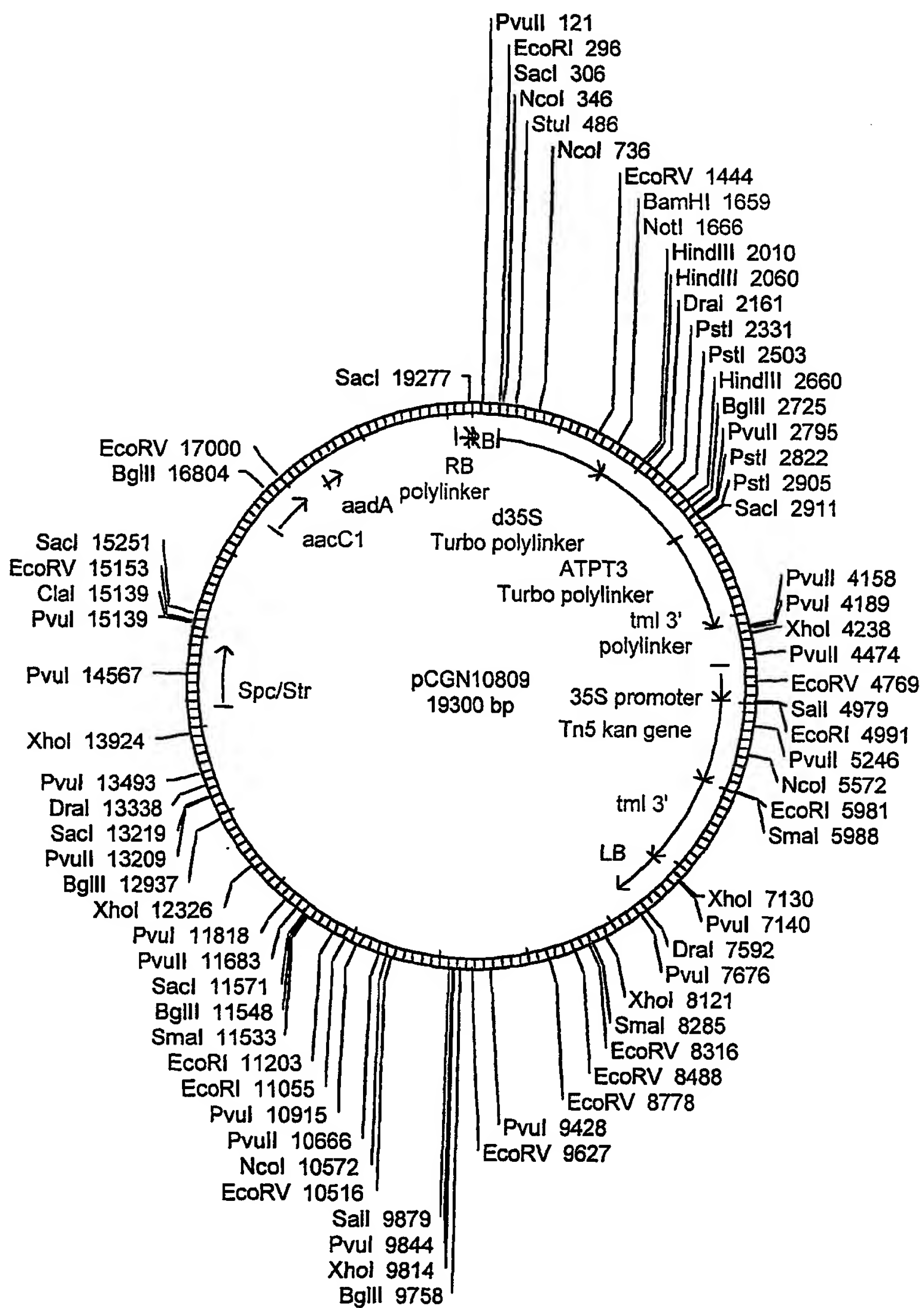


Figure 8

9/40

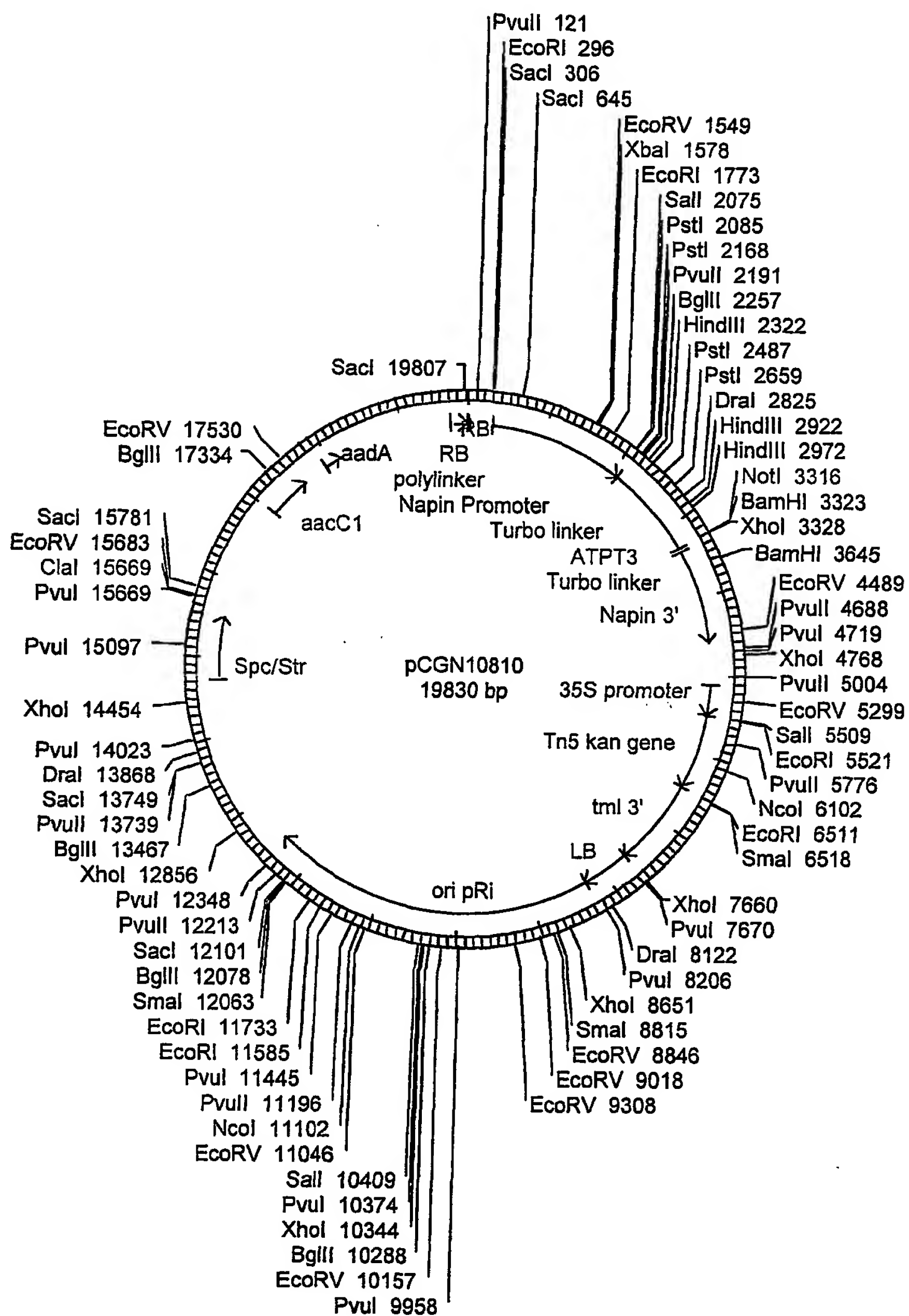


Figure 9

10/40

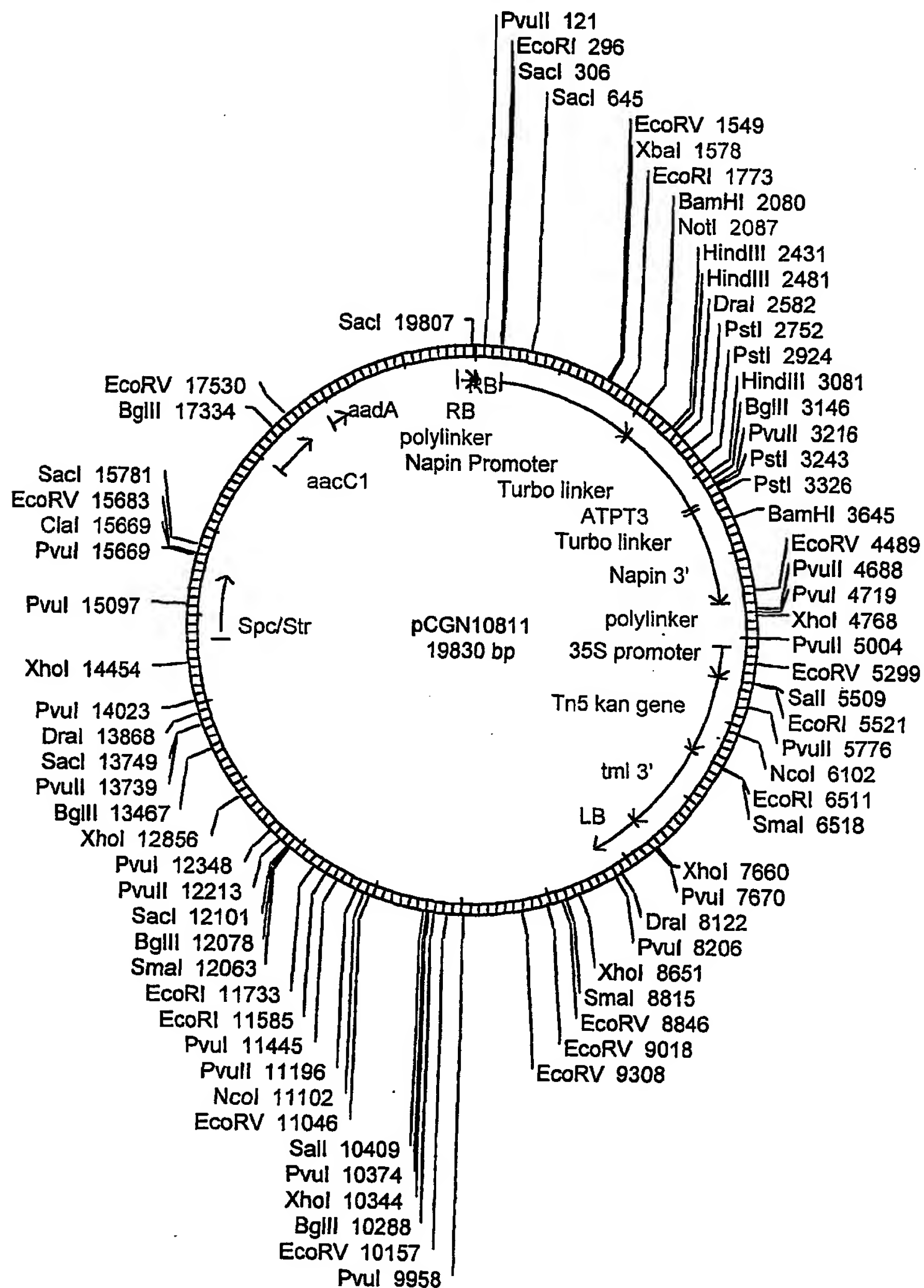


Figure 10

11/40

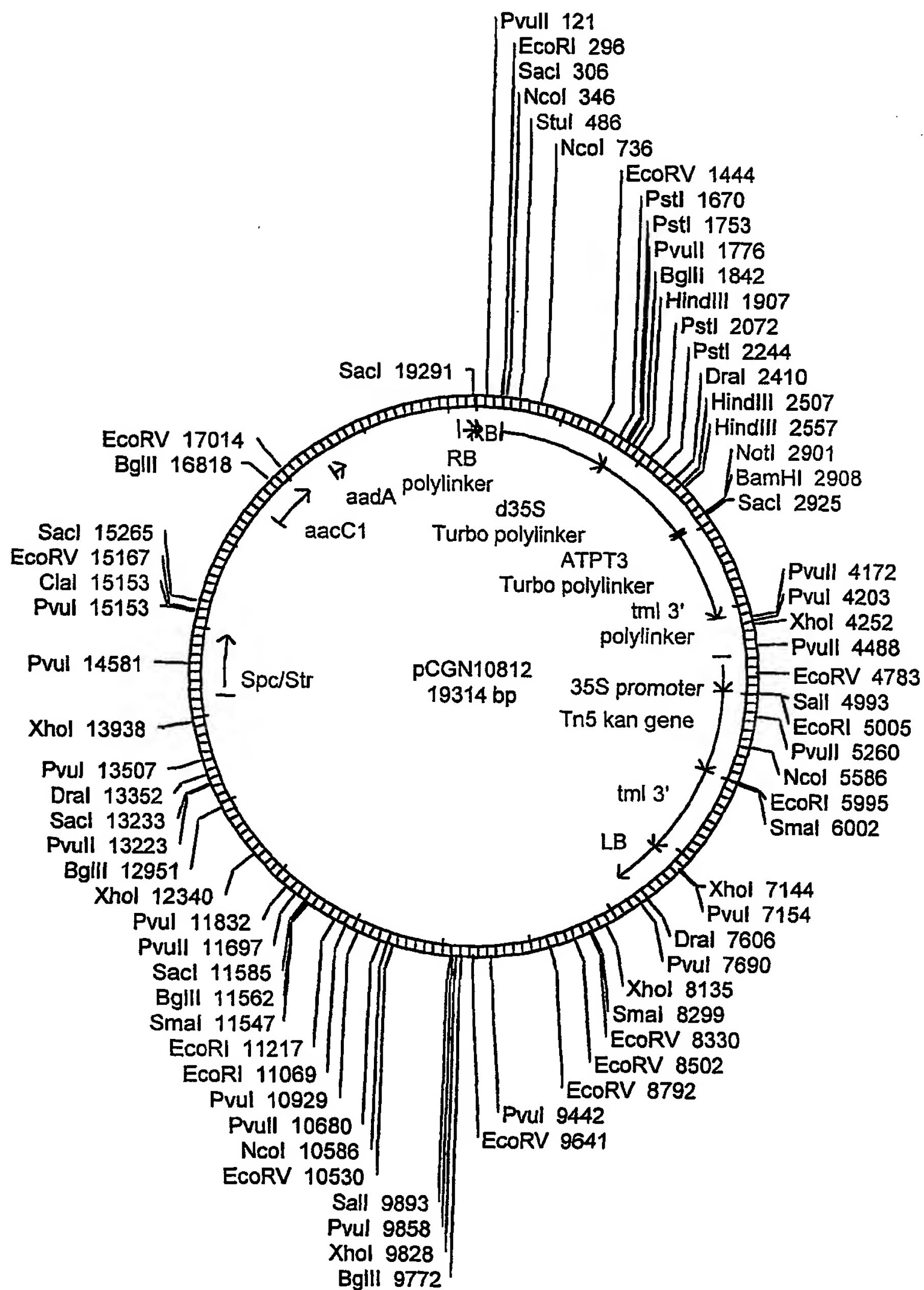


Figure 11

12/40

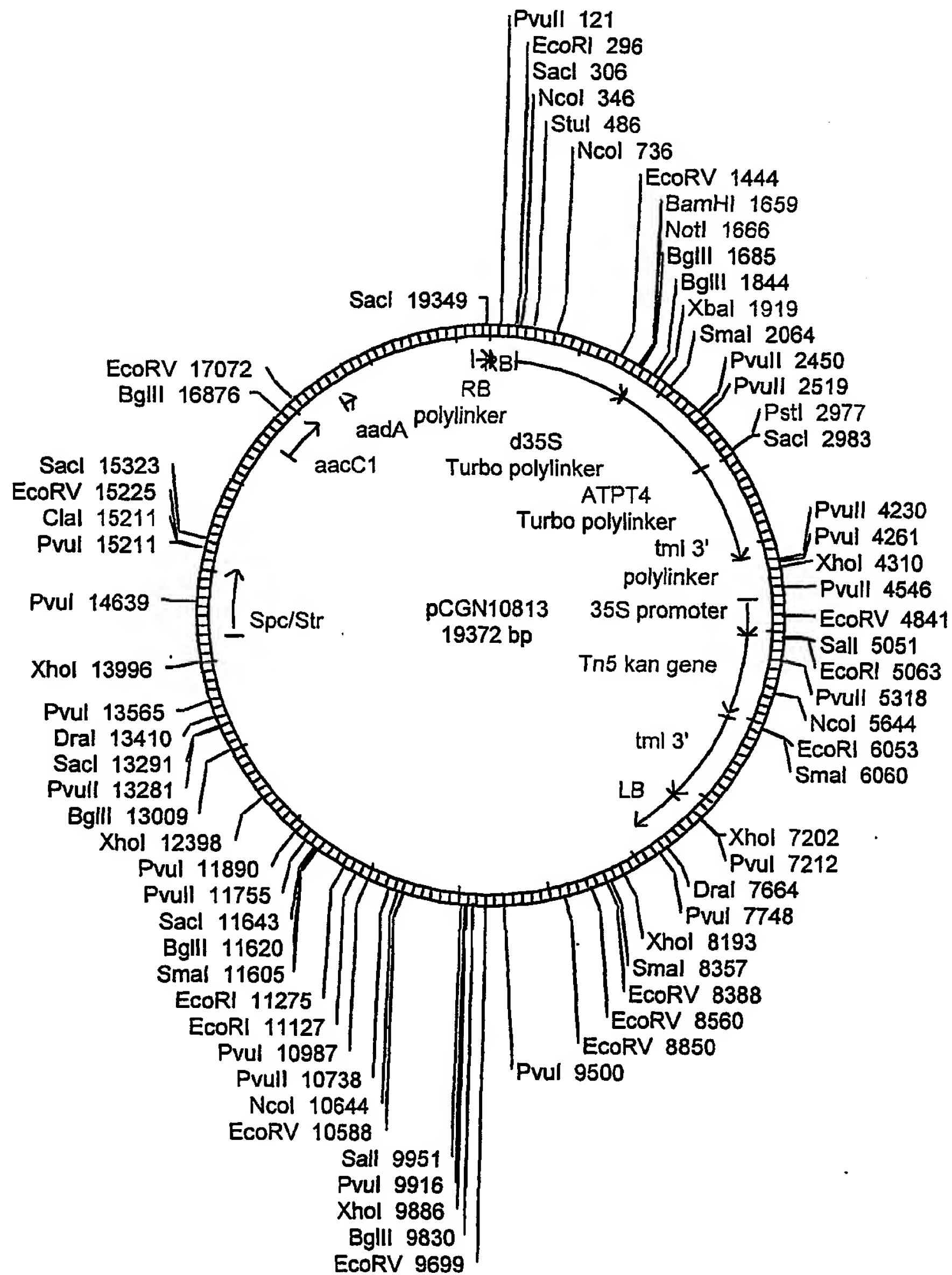


Figure 12



13/40

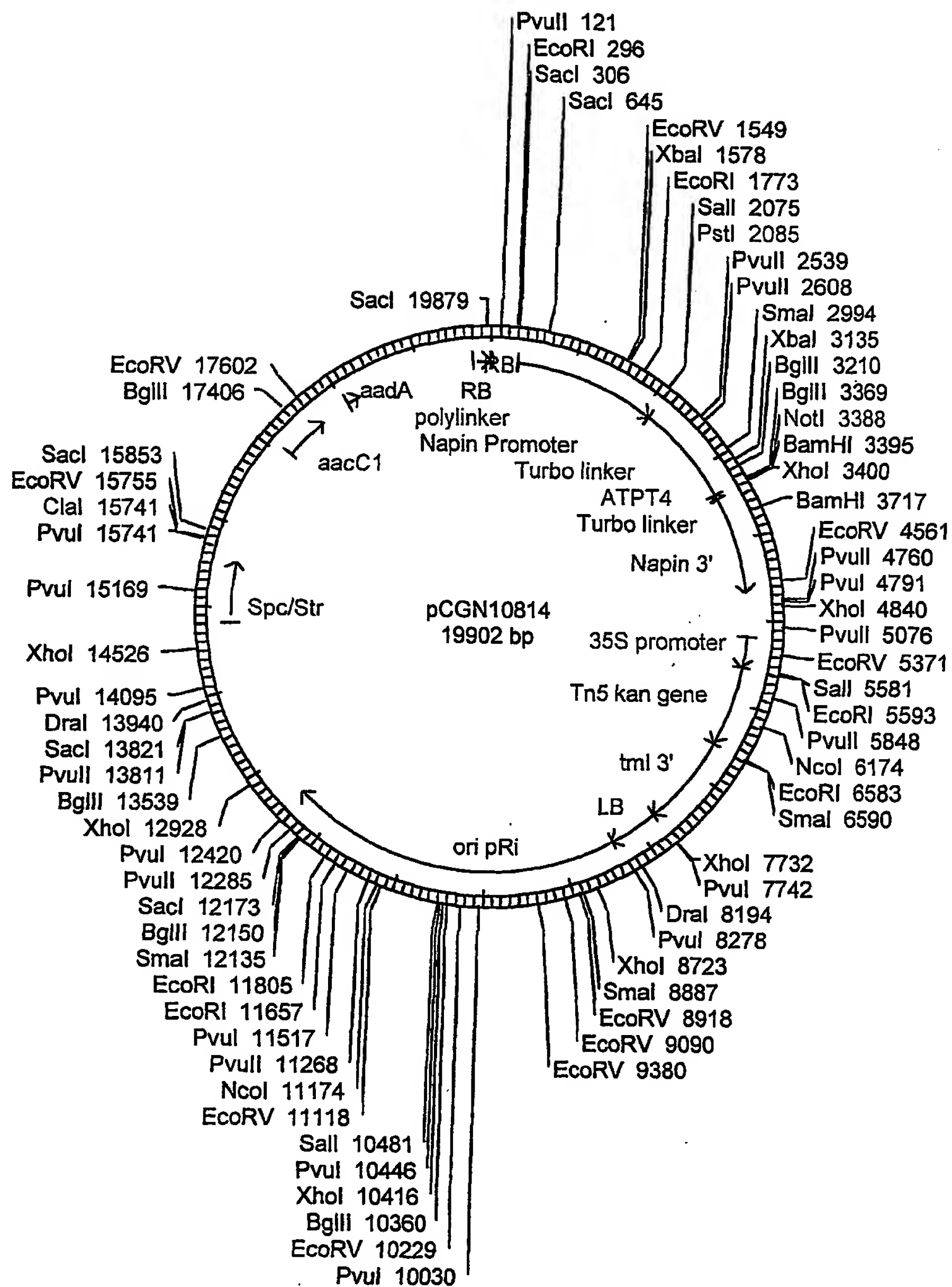


Figure 13

14/40

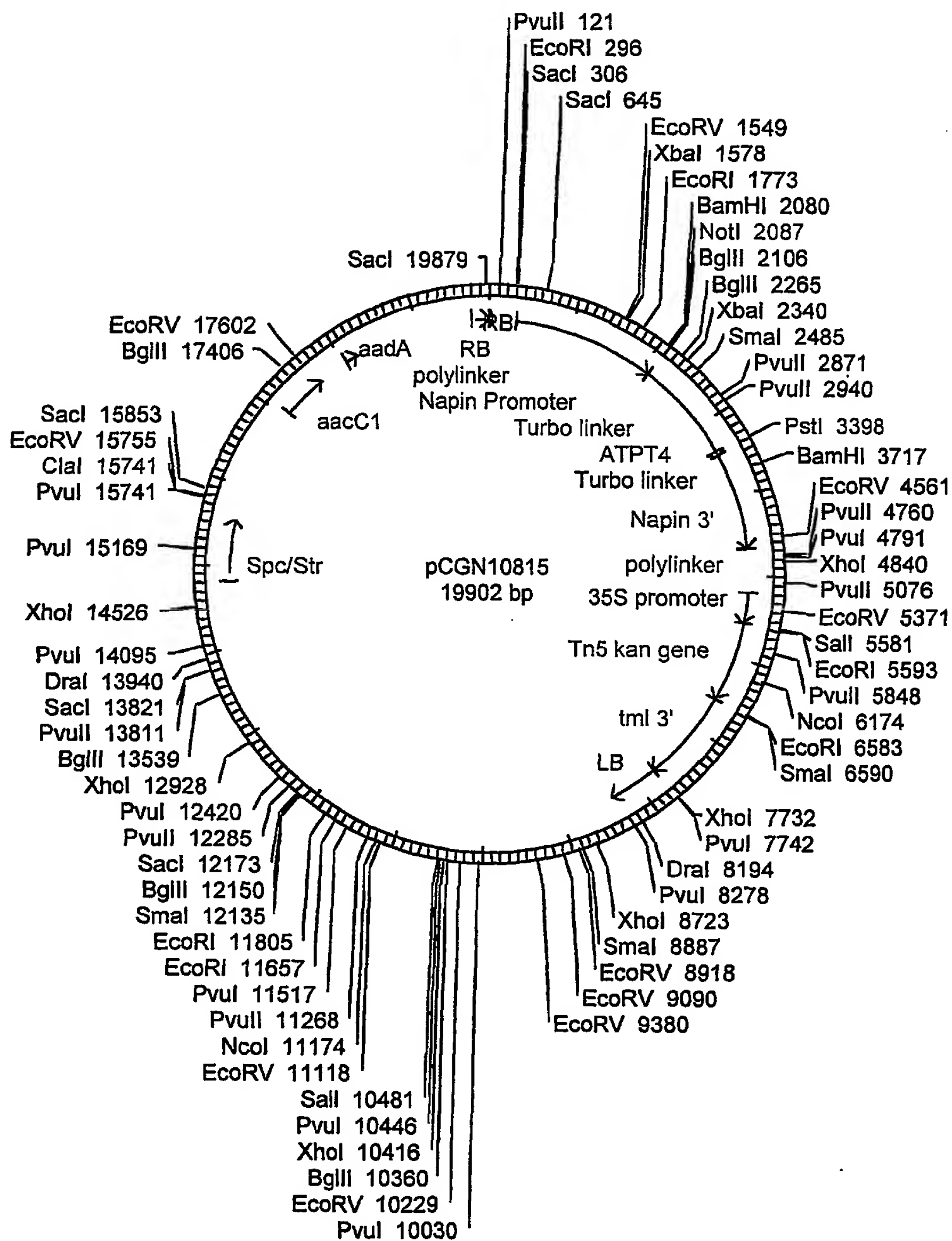


Figure 14

15/40

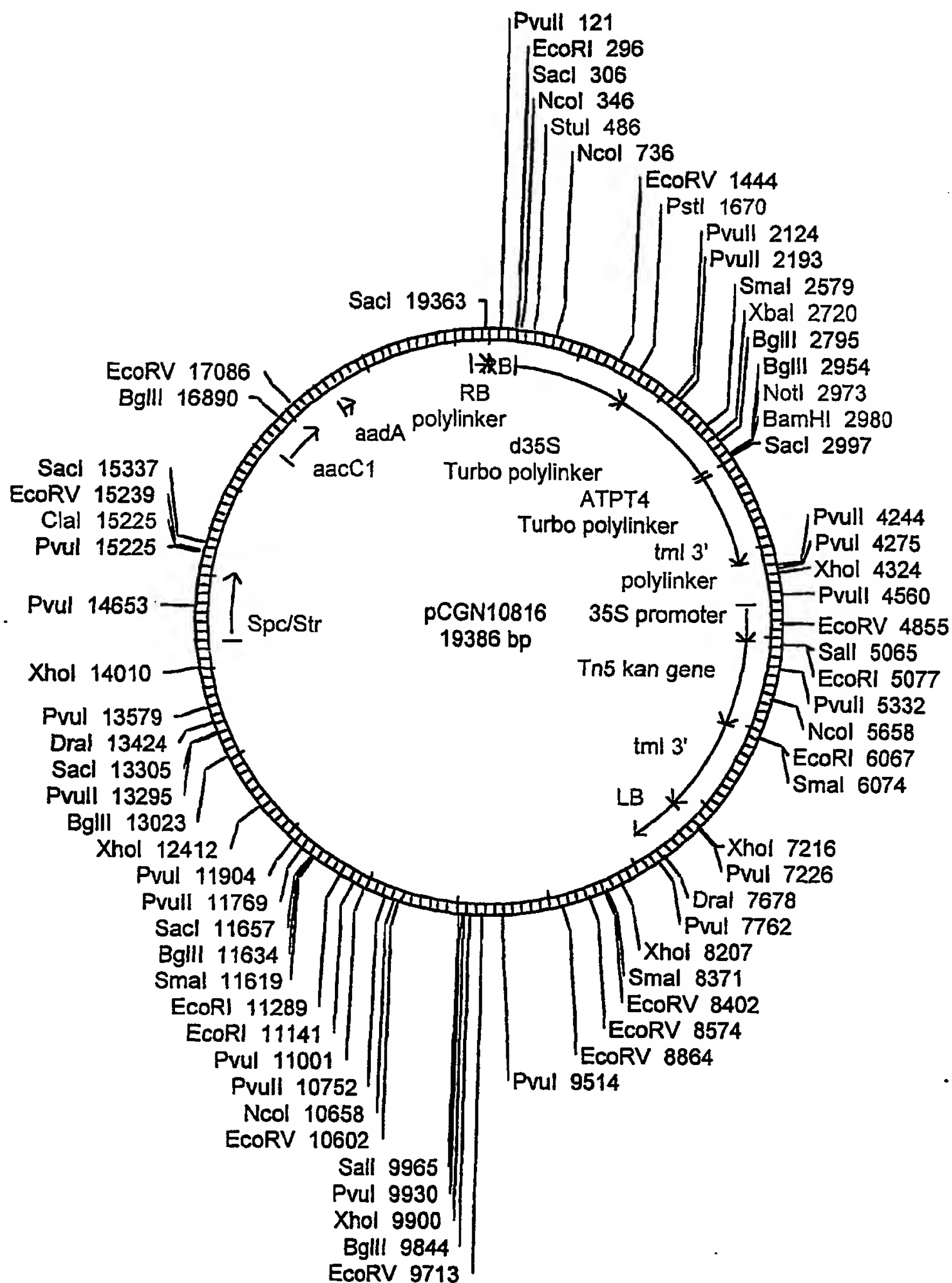


Figure 15

16/40

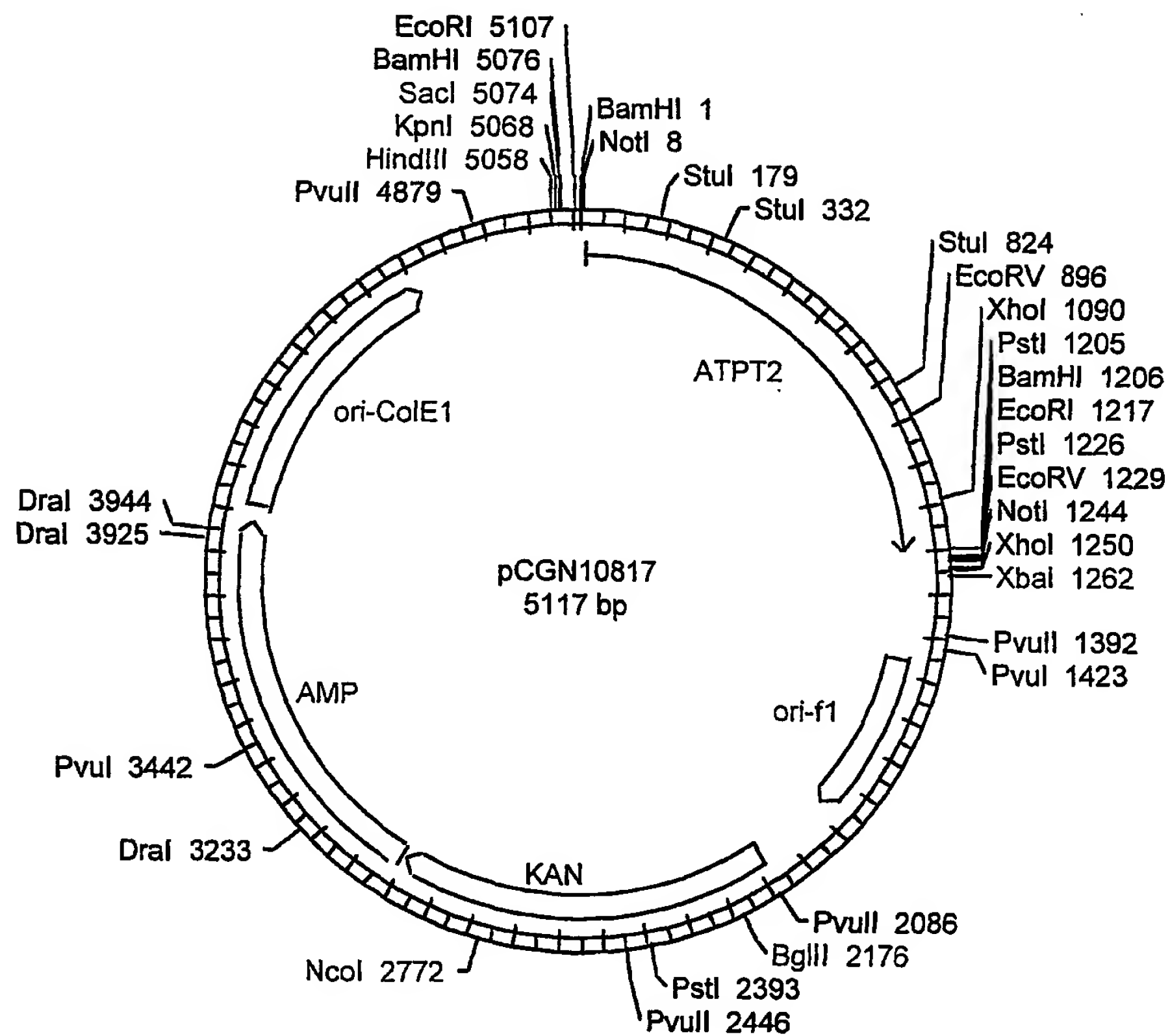


Figure 16

17/40

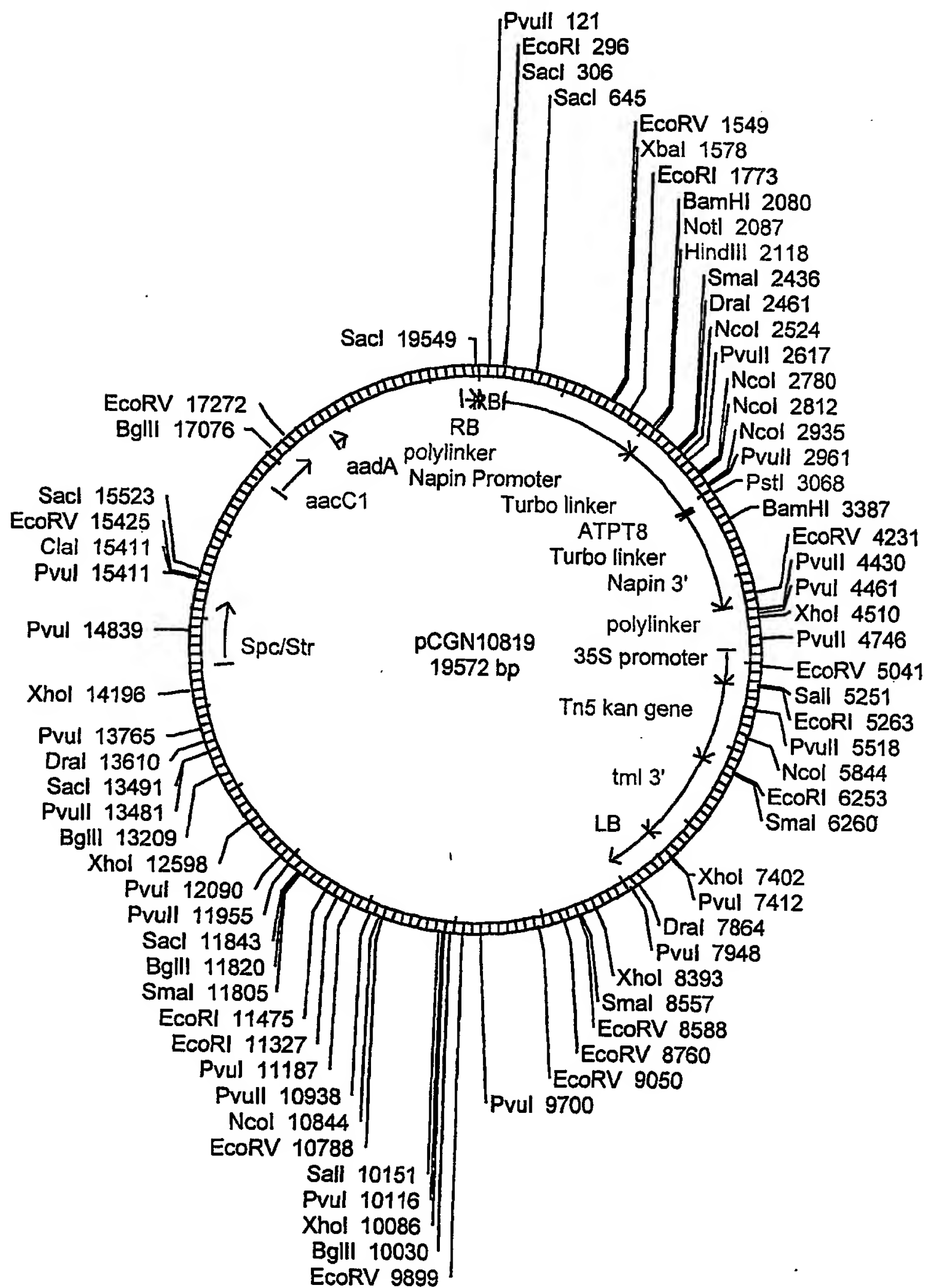


Figure 17

18/40

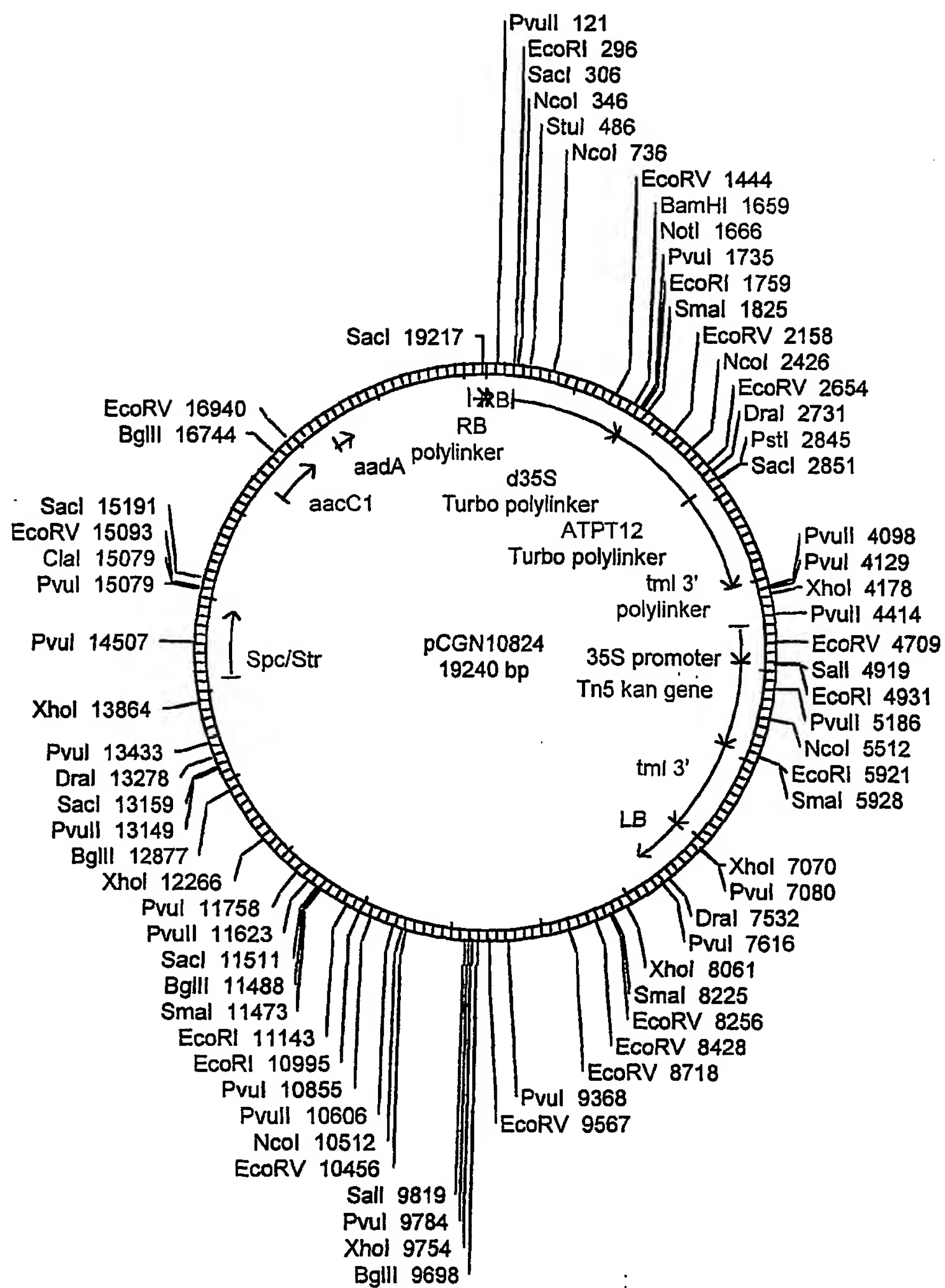


Figure 18

19/40

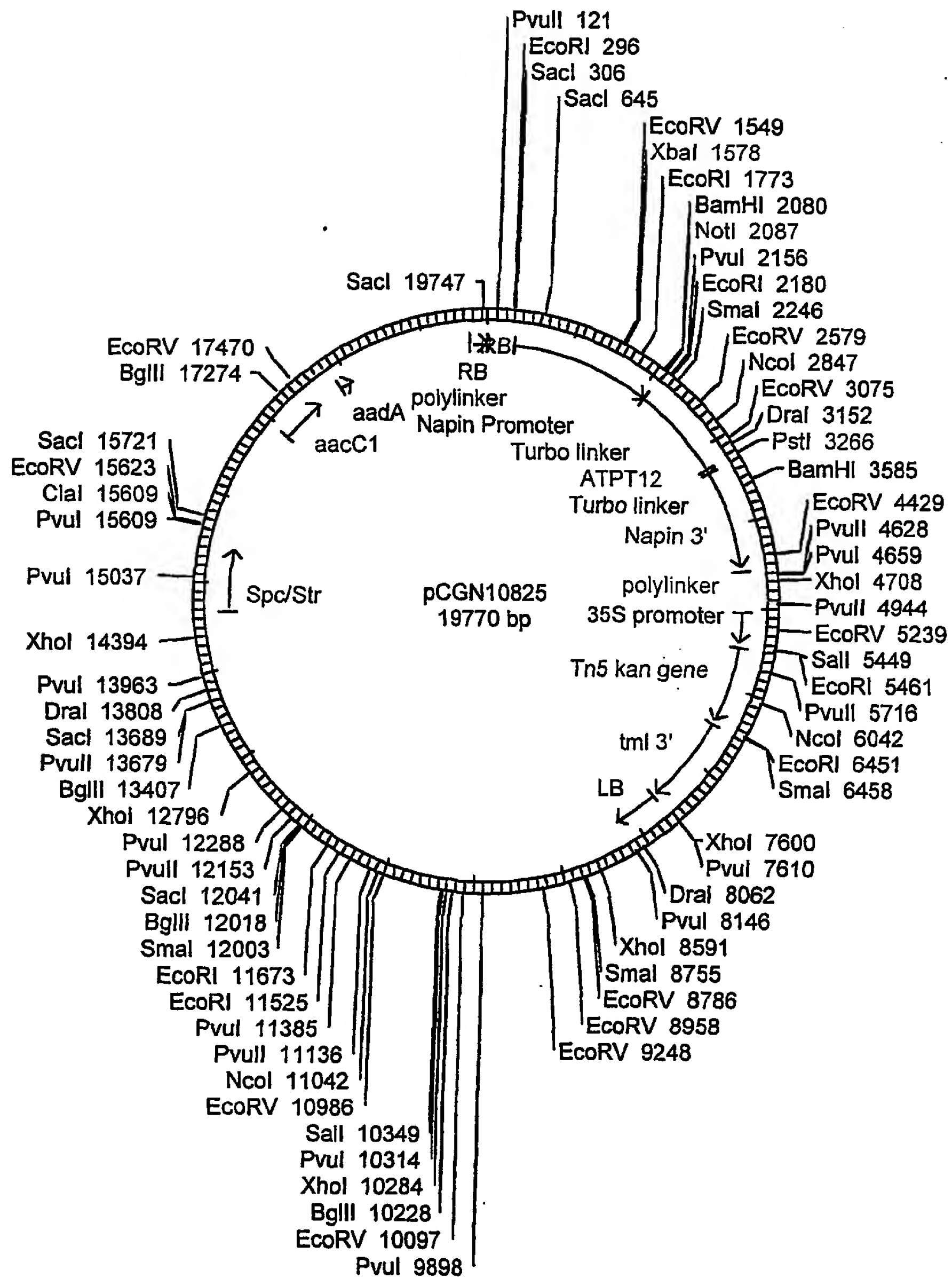


Figure 19



20/40

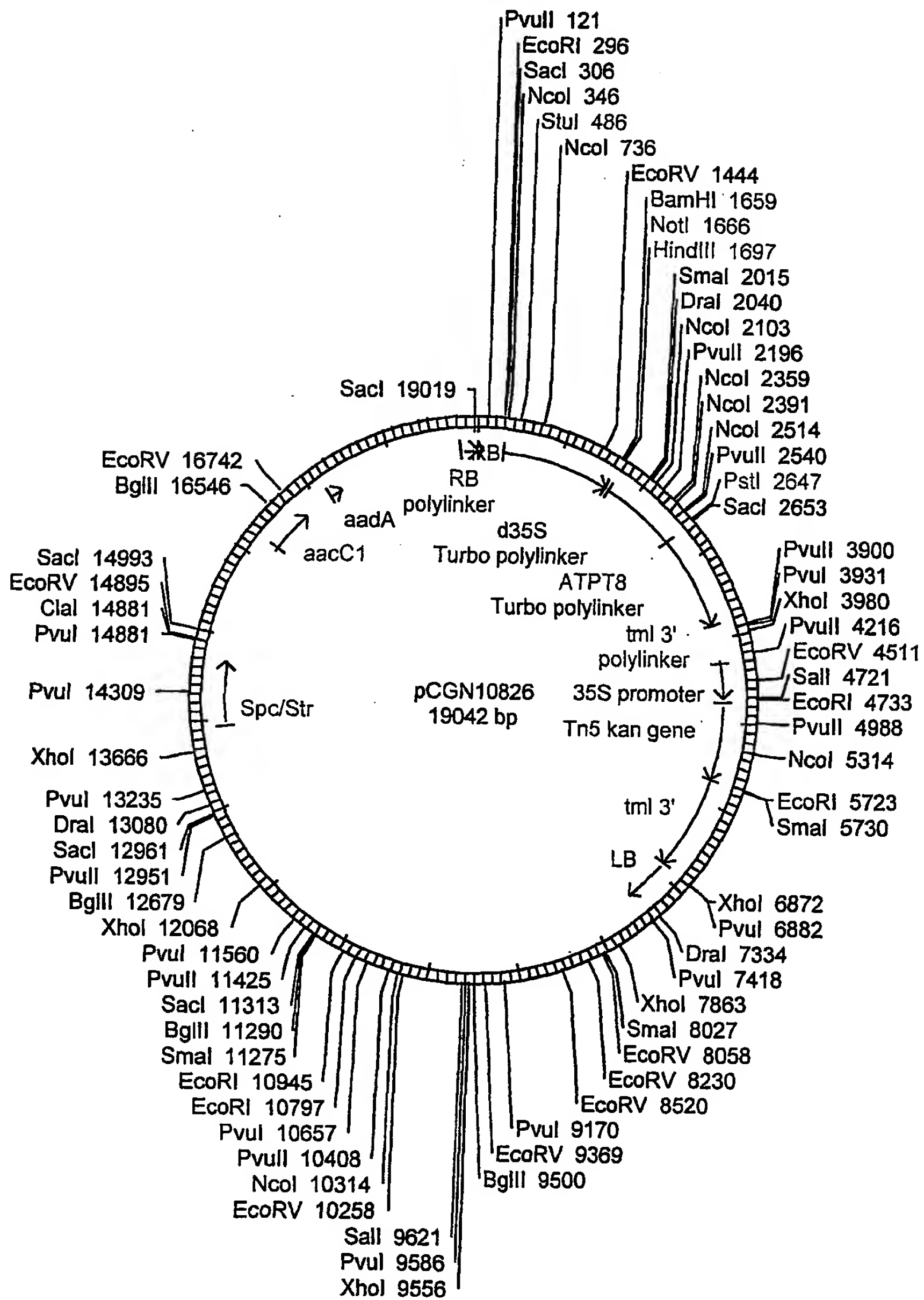


Figure 20

## Figure 21

ATPT2	:	-----	*	20	*	40	*	60	*	80	
SLR1736	:	-----	MESILSSSLVSAAGFCWKKQNLKHLSEIRVLRCDSSKVVAKPKFRNNLVRPDGQSSLLLYPKHKSFRVNATAGQ	:	80						
ATPT3	:	-----	MAFFGLSRVSRRLKSSVSTPSSSALLQSQHKSLSNPVTTHYTNPFTKCYPSWNDNYQVWSKGRLEHQEKFFGVGWNRYRLICGMSSS	:	89						
SLR0926	:	-----	-----		-----						
ATPT4	:	-----	MWRRSVVYRFFSSRISVSSSLPNRPLIPWSRELCAVNSFSQP	:	67						
SLR1899	:	-----	-----		-----						
ATPT12	:	-----	MTSILNTVSTIHSSRVTSDRVGVLSLRNDSVEFT	:	4						
SLR0056	:	-----	-----		-----						
ATPT8	:	-----	-----		-----						
SLR1518	:	-----	-----		-----						

ATPT2	:	*	100	*	120	*	140	*	160	*	1											
SLR1736	:	-----	PEAFDSNSKQK	-----	SFRDS	EDAFYR	-----	FSRPHTV	GTVLS	ILS	-----	VSFLA	VEKVS	-----	DISPL	FTG	ILE	:	140			
ATPT3	:	-----	-----	-----	MAT	IQAFWR	-----	FSRPHTV	GTTL	SMWAVY	-----	LLTIL	GDEN	-----	SVNSP	AS	DLV	FG	:	49		
SLR0926	:	-----	SSVLEGKPKDDKEKSDGVVKKASWIDLYLPEEVRGYAKLARLDKPIGTWELAWPCWMS	-----	-----	IALA	ADPS	-----	LPSEK	YMA	LFGC	:	170									
ATPT4	:	-----	-----	-----	PLW	TIY	-----	LRWHK	PAGRI	ELMIPADWA	-----	VCL	AAQ	-----	G	-----	LPPL	PL	GT	IAL	:	56
SLR1899	:	-----	TAAATATATTG	-----	EISSR	VAALAGLGHYAR	-----	CYWELS	KAKIS	MLVVATSG	-----	TGYIL	GT	-----	NAAIS	FP	GH	CYT	CAG	:	138	
ATPT12	:	-----	TKIHRQHDSMG	-----	AVCK	SYQLTKP	-----	RIIPL	LL	TTAAS	MI	-----	ASEGR	-----	VDLP	KL	IT	ILG	:	60		
SLR0056	:	-----	DKVKSQTPDKAP	-----	AGGS	INQLLGKAS	-----	QETNK	WKIR	QLTKPVT	PPPLV	WGVC	GAAS	-----	ENFHT	WTP	EDY	AKS	ILC	:	139	
ATPT8	:	-----	QNT	-----	GONQ	AKA	-----	ROLL	GMKGAAP	-----	-----	-----	-----	-----	GYIN	SV	ED	ELK	ATC	:	73	
SLR1518	:	-----	EVPKLASAEY	-----	FFKR	GVOGKQF	-----	RSTILL	MA	TALNVRVP	-----	EALIG	EST	-----	DIVT	SE	NR	VRQ	:	63		
	:	-----	SPLAPSTAPAT	-----	RKLW	LAAIKP	-----	PMYT	VA	VPI	TVG	-----	SAV	YGL	ITG	-----	QWHG	DM	ETIF	LL	:	59

ATPT2	:	80	*	200	*	220	*	240	*	260																															
SLR1736	:	AVVAALM	NTIYIVGLNQLSDVE	EDKVN	KPYLPLASGEYS	ANTGIA	IVASFS	IMSFWL	GWIVGS	WPLFWALF	VSF	MLGT	AYS	-----	INTP	PLR	:	228																							
ATPT3	:	AWLACL	IGNVYIVGLNQLSDVE	IDRIN	KPNLPLANGD	ESL	AAQGRW	IVGLCG	VAS	IAI	AWGLG	-----	LWLGL	TVG	LSL	IT	GTAY	-----	SNPP	PVR	:	134																			
SLR0926	:	GALL	-----	IRGAG	CTINDLLDQD	IT	TKVD	RTKLR	PLASGL	IT	PFQ	CGFLG	QOLL	GLG	-----	ILQL	NNY	SRV	LGAS	-----	SL	IVF	:	246																	
ATPT4	:	GTIA	-----	TSGL	GCVNDLWDR	LD	POVE	TKQ	RLA	ARAS	-----	VQVG	AG	VAL	ALL	CAAG	-----	LA	FYLT	PT	SF	WLCVA	-----	AMP	VIV	:	132														
SLR1899	:	TMMI	-----	AASAN	SLNQLFEIS	ND	SKMK	RTMLR	PLPSGR	ESSVPHAN	AWAT	AGAS	GACL	-----	LA	SKTNM	LA	AG	IASAN	-----	LA	LYAF	:	215																	
ATPT12	:	GTIA	-----	AASAQ	TLCIYDQD	ID	YEM	L	TRAR	DI	PAGK	QPRH	AL	GF	LSAL	-----	LA	TFVN	VS	GC	IALSG	-----	IV	FYML	:	137															
SLR0056	:	MMMSG	PCITGYTQT	INDW	DRD	IDA	INEP	YRP	IS	GAISE	PEVIT	QVWV	IT	GG	GI	AGILD	-----	VW	AGHT	TP	TV	FY	ALG	-----	GS	LS	SY	:	223												
ATPT8	:	ML	ESG	PLMTGYTQT	LNDF	YDR	IDA	INEP	YRP	IS	GAISE	PEVIT	QVWV	IT	GG	GI	AGILD	-----	VW	AGHT	TP	TV	FY	ALG	-----	GS	LS	SY	:	157											
SLR1518	:	GTAE	-----	ITEMI	HVASL	HD	VD	LD	DADT	ARGV	GS	LN	VMGN	KMSV	AG	DF	LS	RAC	CAL	-----	AA	LKNT	EV	VAL	LA	TA	VEHL	VT	GETM	:	144										
	:	SAT	IA	-----	UAWIN	LS	ND	VD	FS	ETG	ID	VR	KA	HS	VV	NLT	GN	RNL	VFL	IS	NF	FL	AG	V	L	GL	MS	MS	-----	WRAQ	D	WT	V	LE	IG	VA	-----	FF	EGY	:	138

Figure 22A

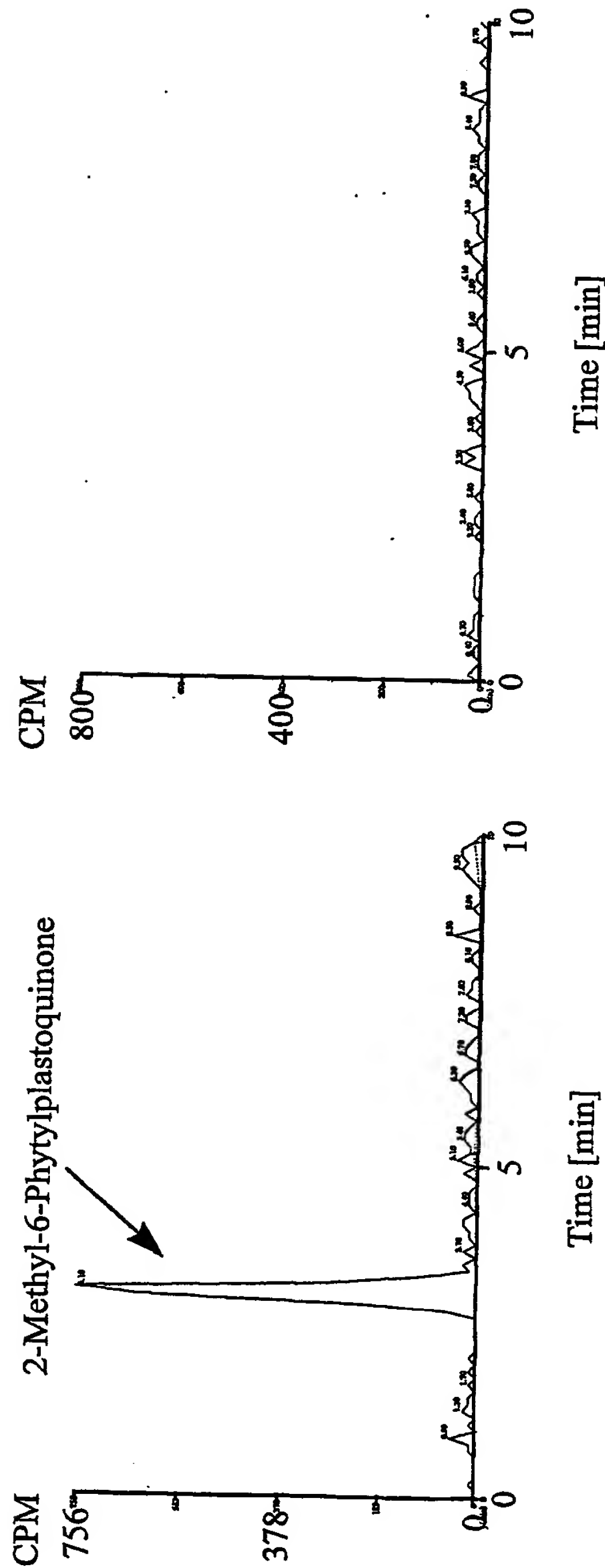
ATPT2 : WKR-FALVAAMCILAARAIINQIAFYH-IQTHVFGRPILFTRPILFATAFMSFFSVVIAIEFKDIPDIEGDKI-FGIRSESVTLG-QKR : 313  
SLR1736 : LKR-FSLLAALCILTVRGIVNGLGLFLF-FRIGLGYPPTLITPFWV-LITLFIILVFTVAIAIEFKDVPDMEGDRQ-FKIQITITLQIG-KQN : 218  
ATPT3 : SMP-LMKRFTFWPQAFGLTINWGAILG-WT--AKGSLAPSLVLPHYLSGVCWTLVYDTIYAHQDKEDBVK-VGVKSTALRFG-DNT : 328  
SLR0926 : AMP-GAKRVFPVQPLVLSIAWGAFLHS--MS--AVTGDLDATWV-LWGAIVFWKIGFDTVYAMADREDERR-IGVNSSALFFG-QYV : 213  
ATPT4 : VMT-PLKQLHPINTWVGAVVGAIPPLIG-WA--WASGOLSYNSMTI-LPAALYFNQIPHFALAHLCRNDYAA-GGYKMTISLFDP--S : 294  
SLR1899 : VVTHWLKRHTAQNIVIGGAAGSIPPLVG-WA--AVTGDLSWTPWV-LFALIFLWPPHFWALALMIKDDYAO-MNMPMPVPIACEKT : 220  
ATPT12 : IYS-APPLKQKQNGWVGNFALGASYISLPWWAGQALFCTITPPDVV-LTLLYSIAGIGIAIVNDFKSVEGDRA-IGLQSTPVAFC-TET : 308  
SLR0056 : IYS-APPLKQKQNGWVGNFALGASYIALPWWAGHAFESTINPTIMV-LTLLYSIAGIGIAIVNDFKSVEGDRA-IGLQSTPVAFC-TET : 242  
ATPT8 : EITSSTEQRYSMDYVMOKTYKKTASITSNCKAVAVLTGQAEVAVLAFEGYGRNLGLAFQIIDDITFTGTSASGCKGSLSDIRH--GV : 231  
SLR1518 : TMOGPPFRIGLGLGELICLITFGPLAI-AAAYYSQSQSFSWNKHT-PSVVFVGISHTAIILFCSHFHQVEDDLA-AGKKSPIVRIG-TKL : 223

ATPT2 : VFWTCVTLIQMAYANAIIVGATSPFINWSKVISVGHVIAATLWAKAKSVDLSSKTEFTS--CYMFIWKIEFYAEYELLPFLK----- : 440  
SLR1736 : VFRGTHILITGCVLAMAIVGFWAAMPLNNTAFETVSHLCILATLWRSRDVHLESKTEFTS--FYQFIWKIEFYAEYELLPFLK----- : 393  
ATPT3 : KLWITGFGTASIGFIALSGFSADLGWQYASIAAASGQGWQIGTADLSSGADCS-----RKEVSNKWPFGATFSGVVLGRSFQ- : 407  
SLR0926 : GEAMGTEFFALTIGCLFYLGNTLMLNPLYWLSIAHAT--GWNVHQYIQLSAPTEP-KFY-----GQIFGQNMIIIGFVILLAGMLLWGL-- : 292  
ATPT4 : GKRFAAVALRNCFYMIPLGFETAYDWGLTSSWFCTESTITLALIAATAFSFYRDRTHKA-----RKMTHASILELPVFNMSGILLHVRVSN : 379  
SLR1899 : VSQIWIYSSLLVVPFSLLVYPLHQLGILYLADAHIT--GGQFELVKAWQLKQAPGDRDIA-----RGLFKFSHFYLMILICLAMVDDSLPVT : 303  
ATPT12 : AKWICWGAIDITQLSVAGYILASCKPYVALALVALH--TPQVVFQFKYFLKDPVKYDK-----YQASAPPELVLFGEVTAIASQH-- : 387  
SLR0056 : AAWICMIMIDVFOAGIAGYILVYHQQLYATINBLLH--TPQVVFQFKYFLKDPVKYDK-----YQASAPPELVLFGEVTAIASQH-- : 324  
ATPT8 : ITAPIIFAMEEFQUREVVDQVEKDPNRNVDIANEYTGKSKGTQARARELAMAHANLAAAIGSLPETDNEDKRSRRACIDITHRVITRN : 320  
SLR1518 : GSQVLTLSVVSLLYLTATAGVICHQAPWQTLTASHPWAVQVLRHVQYHQDQPEQVSNCK--FIAVNLHFFSGMLMAAGYGWAGLG-- : 307

ATPT2 : ----- :  
SLR1736 : NTIF----- : 308  
ATPT3 : ----- :  
SLR0926 : ----- :  
ATPT4 : NQQQLVEEAGLTNSVSGEVKTQRRKKRVAQPPVAYASAAPFPLPAPSFYSP : 431  
SLR1899 : --HQLVAQMGTLLLG----- : 316  
ATPT12 : ----- :  
SLR0056 : ----- :  
ATPT8 : K----- : 321  
SLR1518 : ----- :

Figure 22B

24/40



*Synechocystis* 6803 wild type      *Synechocystis* slr1736 knockout

Figure 23



25/40

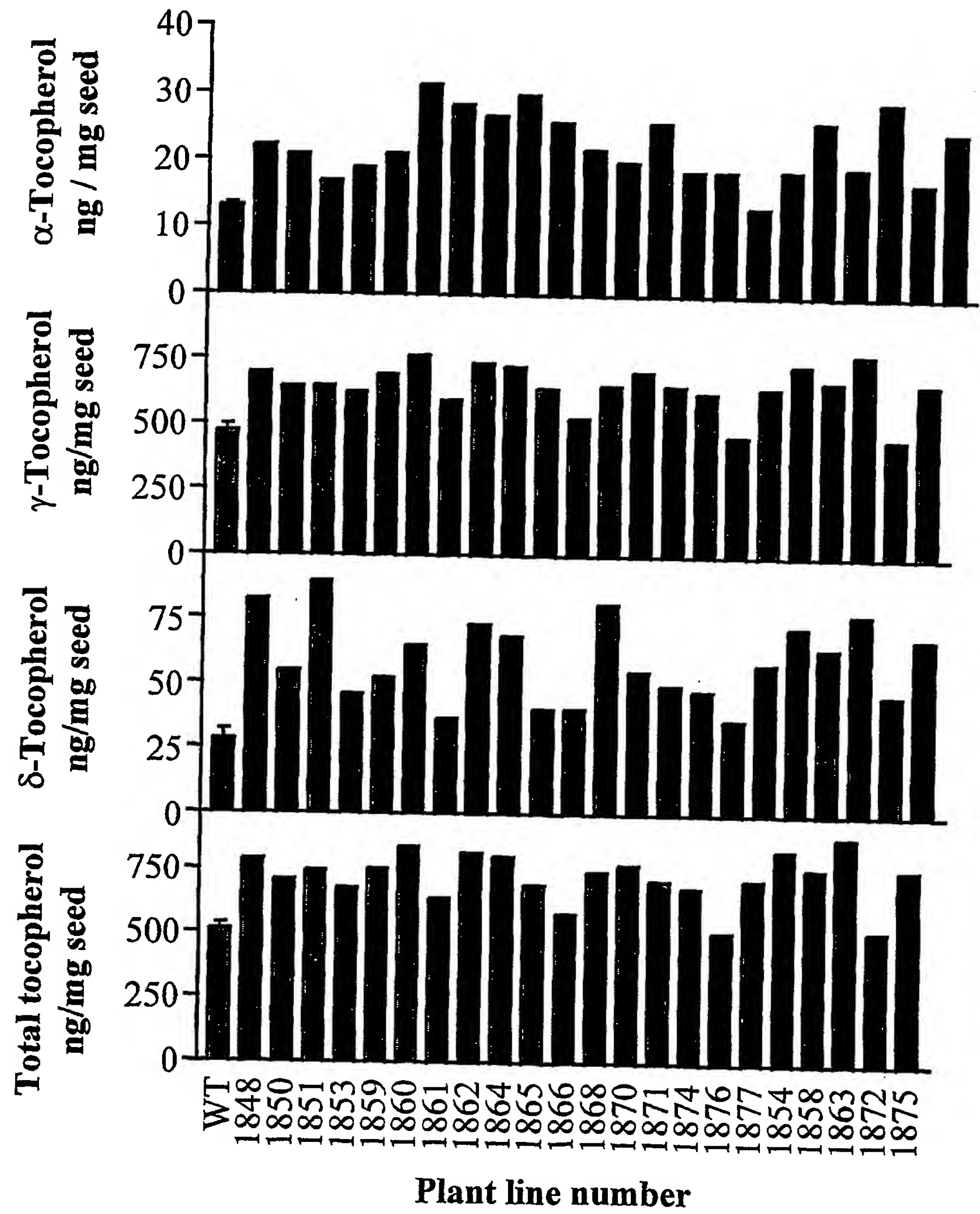


Figure 24

26/40

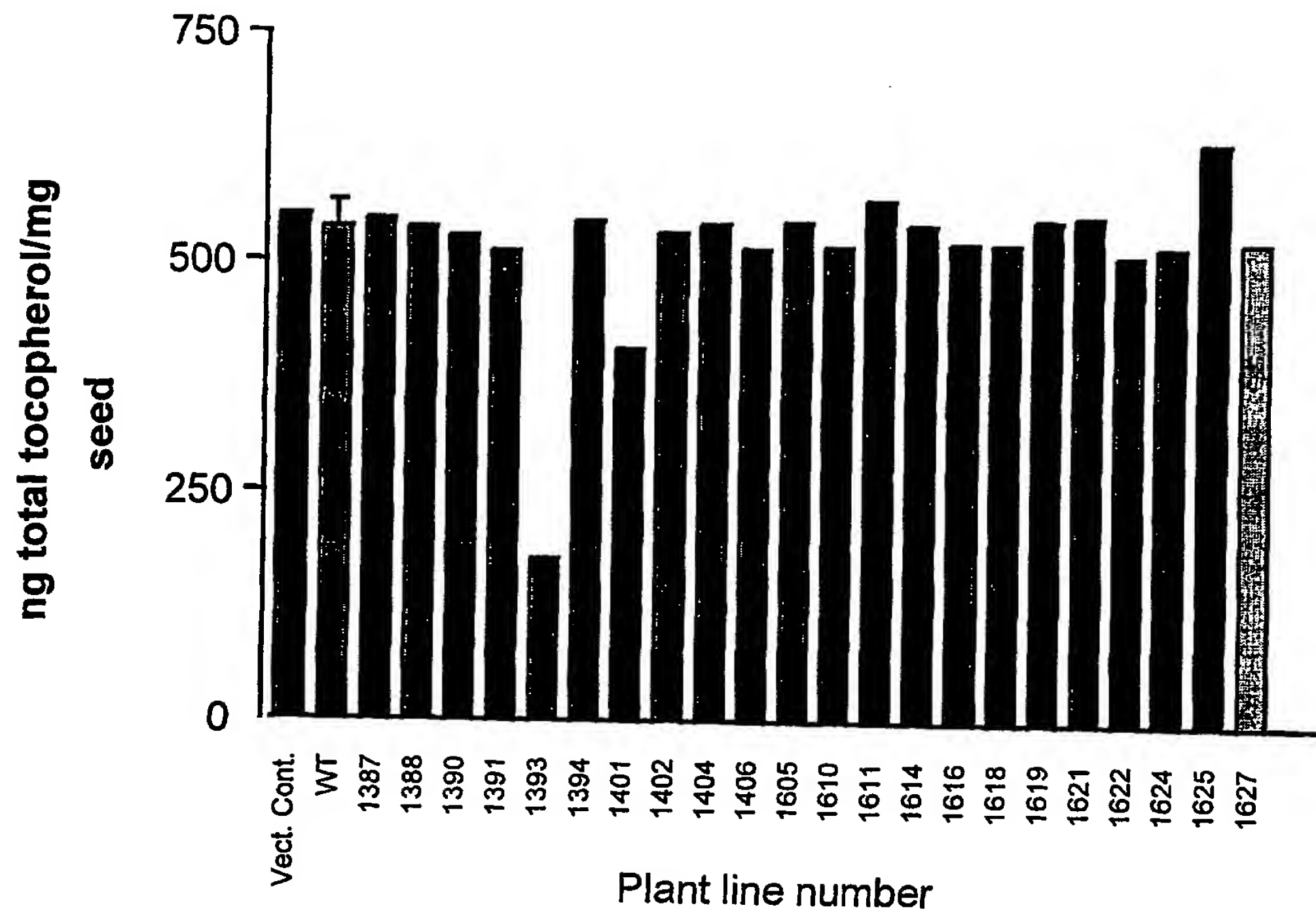


Figure 25



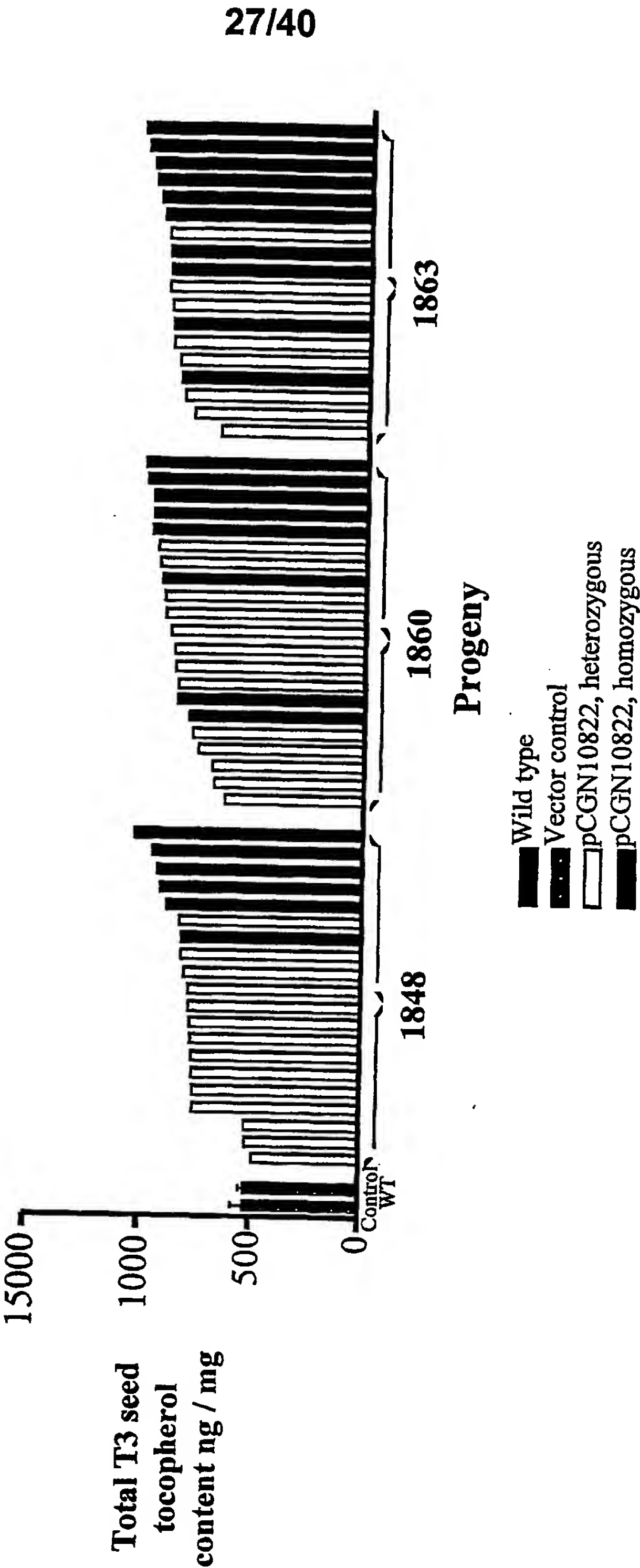


Figure 26

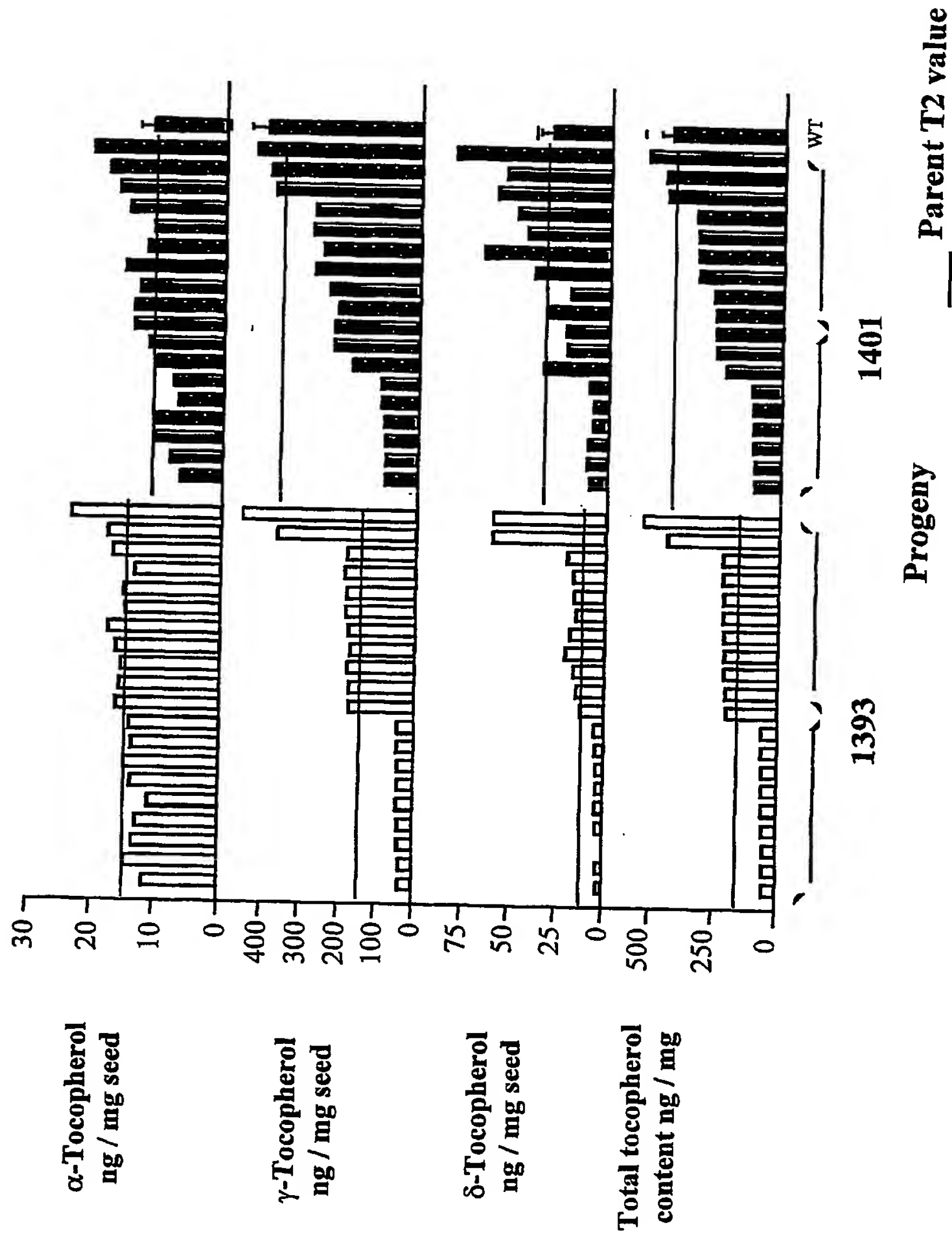
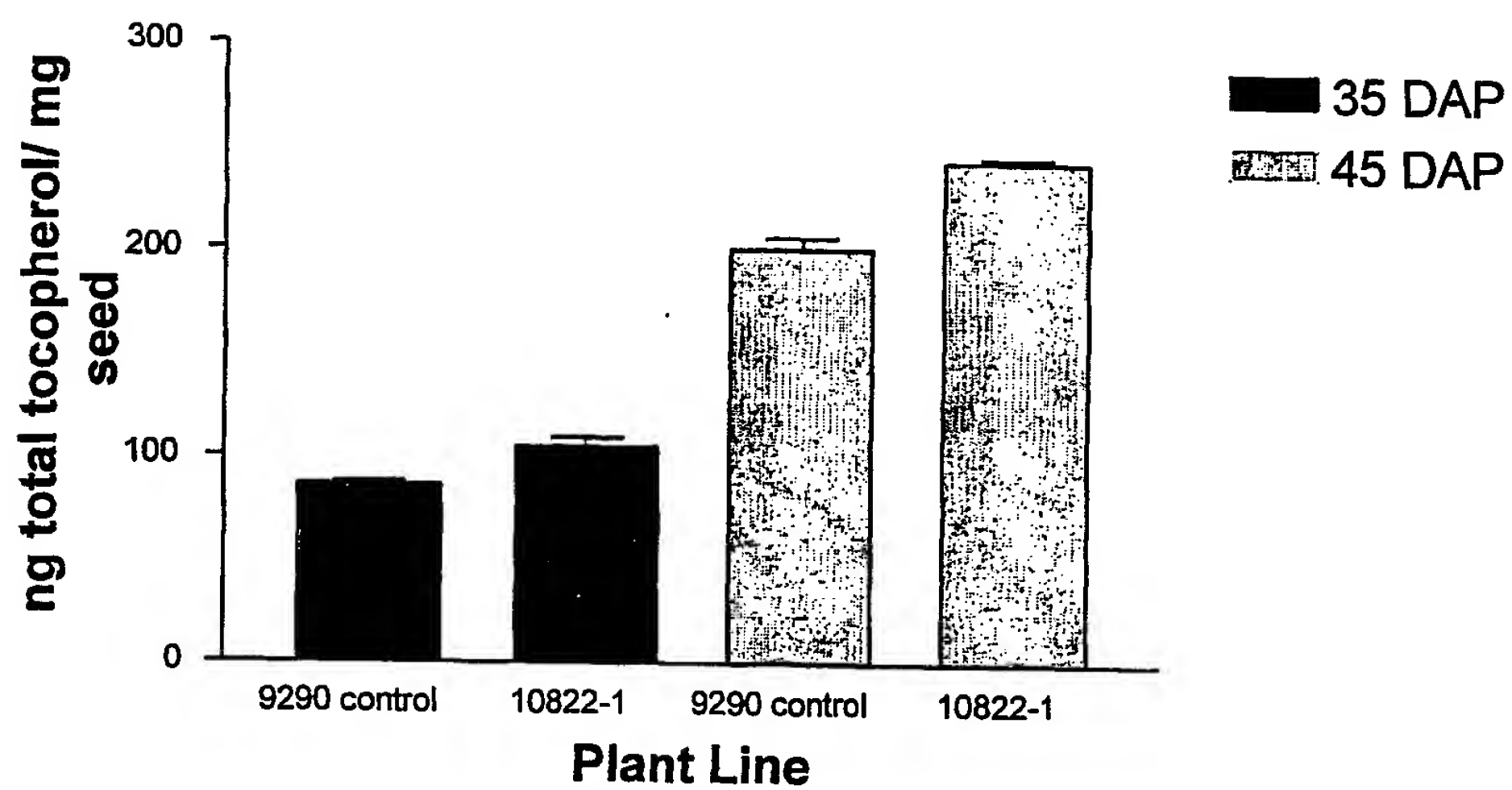


Figure 27

29/40

### Total tocopherol in Napin ATPT2 Canola Seed

**Figure 28**

30/40

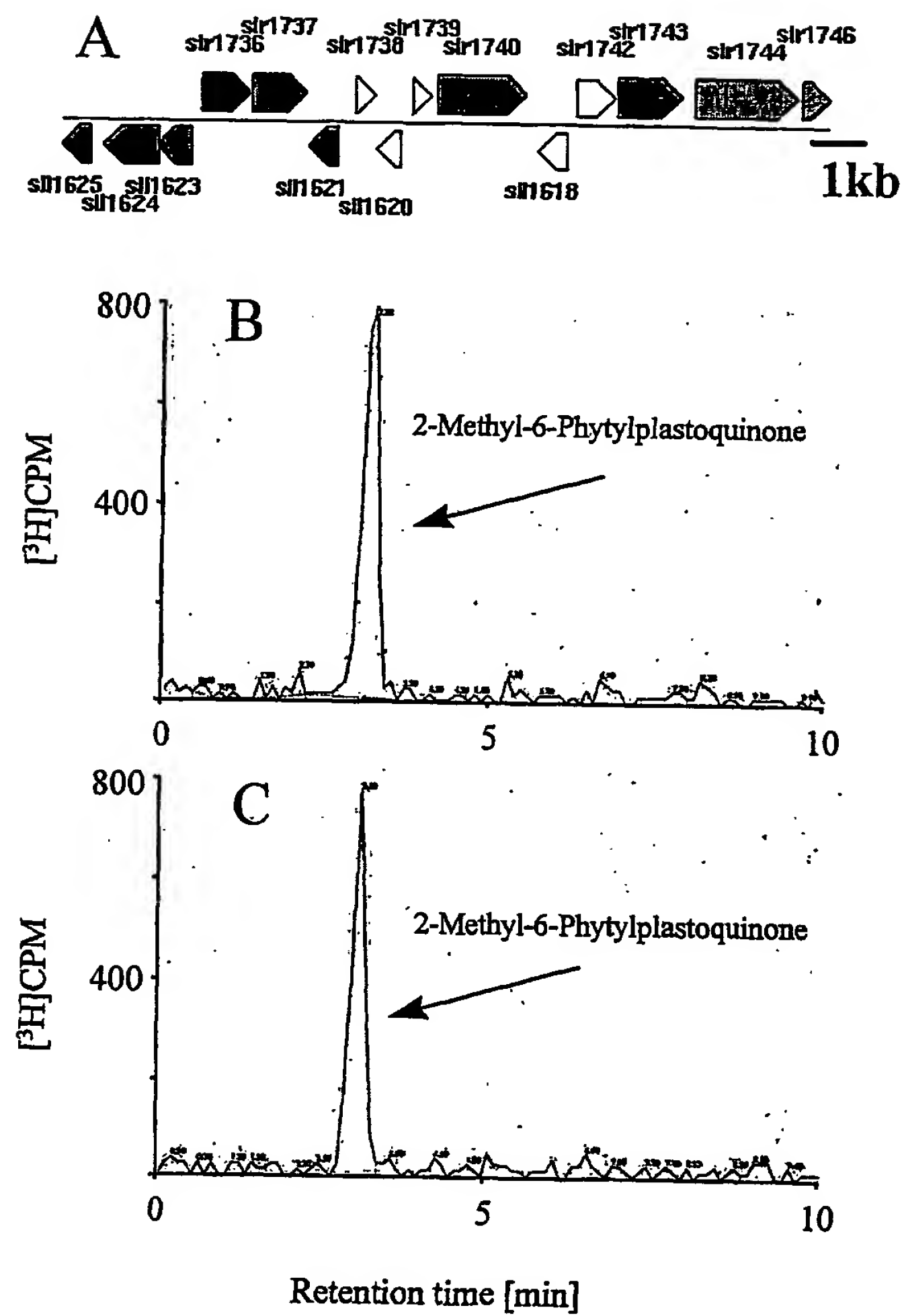


Figure 29

31/40

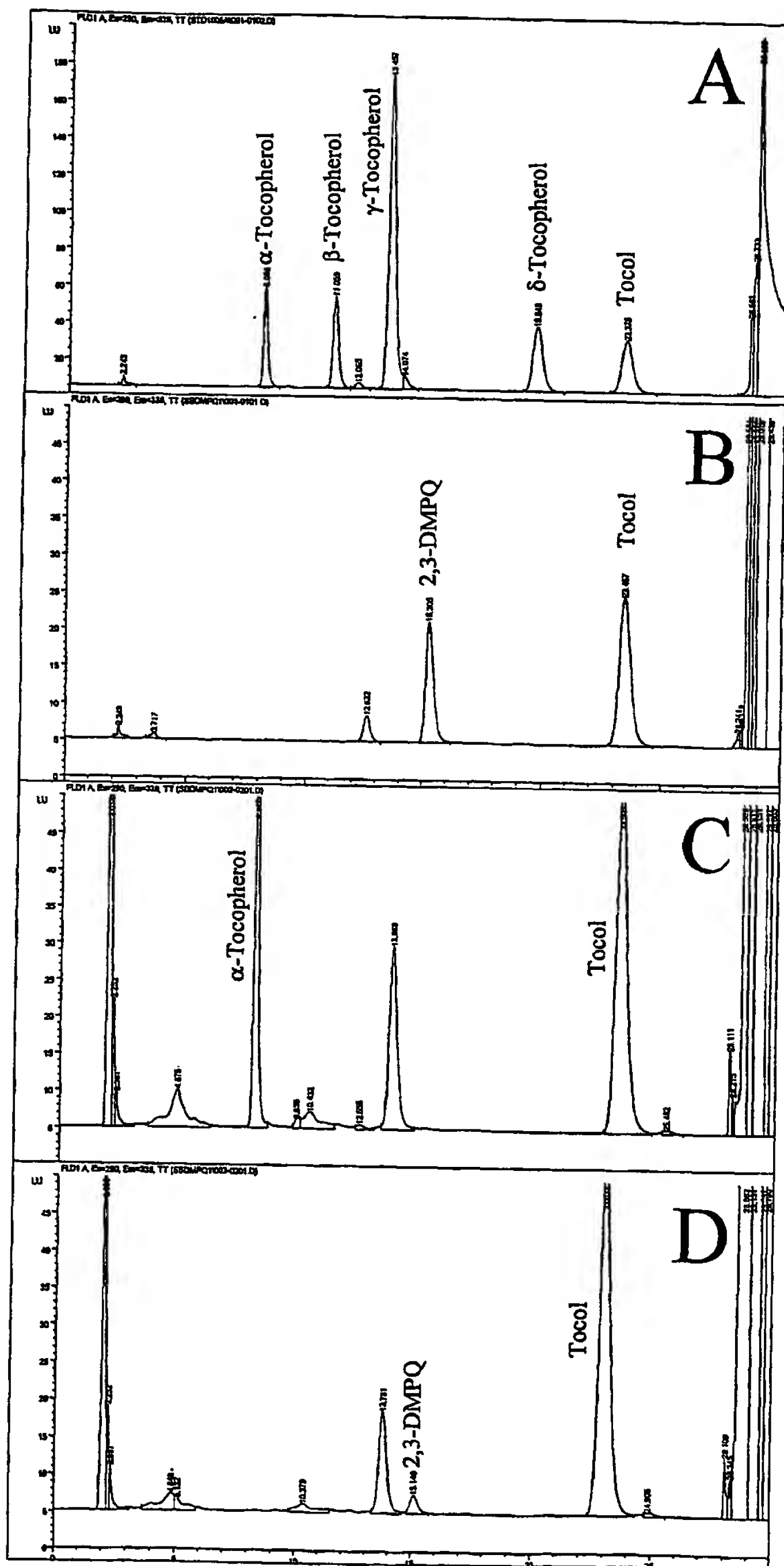


Figure 30

32/40

Query Sequence: F4D11 AL022537  
Database: PIR\_T04448.atcea.list.fasta  
Database: PIR\_T04448  
Plus (+) denotes forward strand, and minus (-) reverse strand.  
Asterisks (\*) denote bases not shown on pair wise alignments.

Alignment 1

Query- genomic 12194 CACACGTTCTCGTCCCTTTCTTCTTCTCTCTGCATTCTTCACAGAGTTTGTCAACCA  
ATCEA4C371+ 1 ----- C. est  
Met  
Query- 12134 ~~ATGGA~~ ATGGA  
ATCEA4C371+ 2 ACCCCAAACATCACAAATTCACATTCTTTTGCATATTTCTTCTTCTTCTCCATTATGGA  
Query- 12075 GATACGGAGCTTGATTGTTTCTATGAACCCTAATTTATCTTCTTCTGAGCTCTCTCGCCC  
ATCEA4C371+ 62 GATACGGAGCTTGATTGTTTCTATGAACCCTAATTTATCTTCTTCTGAGCTCTCTCGCCC  
Query- 12015 TGTATCTCCTCTCACTCGCTCACTAGTTCCGTTCCGATCGACTAAACTAGTTCCCGCTC  
ATCEA4C371+ 122 TGTATCTCCTCTCACTCGCTCACTAGTTCCGTTCCGATCGACTAAACTAGTTCCCGCTC  
Query- 11955 CATTTCTAGGGTTTCG ~~ATCTCCACCCGAATAGTGAACTGACAAGATCTCCGT~~  
ATCEA4C371+ 182 CATTTCTAGGGTTTCGCGCTCGATCTCCACCCGAATAGTGAACTGACAAGATCTCCGT  
Query- 11895 TAAACCTGTTTACGTCCCGACGTCTCCCAATCGCGAACTCCGGACTCTCACAGTGGGTA  
ATCEA4C371+ 242 TAAACCTGTTTACGTCCCGACGTCTCCCAATCGCGAACTCCGGACTCTCACAGTGG  
here Synecho seq aligns from  
Query- 11835 AATTGATCCATTCCATTCTCTTCTTCTTCTTGTGTTTATTAAAGCTCCAATTCAG  
ATCEA4C371+ 299  
~~ 60 bp removed ~~  
Query- 11715 \*\*\*\*\*TTTG  
ATCEA4C371+ 299  
PIR:T04448 1  
Query- 11655 GTGGCTCACCATTGACGACTACTTTTGAATTTGAGTTTTTGAAAAATGCAATTAACAT  
ATCEA4C371+ 299  
PIR:T04448 1  
M Q F N I  
arab sequence which is incorrect

Figure 31A

33/40

Query- 11595 CAGAGAGTTTTTTTATGGTTGATAACTTATTGTTTAACTTTTGAAAAATGCAGATA  
ATCEA4C371+ 299  
PIR:T04448 6 R E F F F L W L I T Y C L T F E K C R Y

Query- 11535 CCATTTCGATGGAACACCTCGGAAGTCTTCGAGGGATGGTATTTTCAGGGTTTCCATCCC  
ATCEA4C371+ 302 CCATTTCGATGGAACACCTCGGAAGTCTTCGAGGGATGGTATTTCTCCATCCC  
PIR:T04448 26 H F D G T P R K F F E G W Y F S I P

Query- 11475 AGAGAAGAGGGAGAGTTTTTGTATTATGTATTCTGTGGAGAATCCTGCATTTCCGCAGAG  
ATCEA4C371+ 362 AGAGAAGAGGGAGAGTTTTTGTATTATGTATTCTGTGGAGAATCCTGCATTTCCGCAGAG  
PIR:T04448 46 E K R E S F C F M Y S V E N P A F R Q S

Query- 11415 TTTGTCACCATTGGAAGTGGCTCTATATGGACCTAGATTCACTGGTGTGGAGCTCAGAT  
ATCEA4C371+ 422 TTTGTCACCATTGGAAGTGGCTCTATATGGACCTAGATTCACTGGTGTGGAGCTCAGAT  
PIR:T04448 66 L S P L E V A L Y G P R F T G V G A Q I

Query- 11355 TCTTGCGCTAATGATAAATATTTATGCCAATACGAACAAGACTCTCACAATTTCTGGGG  
ATCEA4C371+ 482 TCTTGCGCTAATGATAAATATTTATGCCAATACGAACAAGACTCTCACAATTTCT  
PIR:T04448 86 L G A N D K Y L C Q Y E Q D S H N F W G  
ATCEA4C371+ Exon 11538 11301 Confidence: 100 100

Query- 11295 AGGTAACCTCCTTGACCCTTAAATGCTGTGTCATGACAATAAGAAATCATATCTGAGTCT  
ATCEA4C371+ 537  
PIR:T04448 106 D  
PIR:T04448 Exon 11609 11294 Confidence: 100 100

Query- 11235 TTTCTCTACTTCTAGTACTAATGTTTCGTTATTGTTGTTAAAGATCTAAGTCTTATCTGAA  
PIR:T04448 107

Query- 11175 TTTTGTTACATTTTGGTTCTGGTGCTTTCTCAACATGAATTTGTATATATGACTTTAAAG  
PIR:T04448 107

Query- 11115 ATTGCTTACCTAAAGTTTTTACTCATGCATAGATCGACATGAGCTAGTTTTGGGGAATAC  
PIR:T04448 107 R H E L V L G N T

Query- 11055 TTTTAGTGCTGTGCCAGGCGCAAAGGCTCCAAACAAGGAGGTTCCACCAGAGGTTCTCAC  
PIR:T04448 116 F S A V P G A K A P N K E V P P E  
PIR:T04448 Exon 11083 11004 Confidence: 96 100

Figure 31B



34/40

Query- 10995 TCCTCCCTTGTGGTTACTTTGTTATCTGTTAAATAGTTTTCCAATTGTATCCGGATAGT  
PIR:T04448 133

Query- 10935 GTTCTACTTCTCCTTGTAGAAAATCTCAAGTTTTGTTACTCTTGCTATTCTCTTGGATG  
PIR:T04448 133

Query- 10875 TTGATTTGTAAAGCATGTCGTTTATTGTAGGAATTTAACAGAAGAGTGTCCGAAGGGTT  
PIR:T04448 133  
E F N R R V S E G F

Query- 10815 CCAAGCTACTCCATTTTGGCATCAAGGTCACATTTGCGATGATGGCCGGTAATTATATGA  
PIR:T04448 143  
PIR:T04448 143  
Exon 10844 10768 Confidence: 100 100  
Q A T P F W H Q G H I C D D G R

Query- 10755 TTCTATGCACAACAAGAATTCATATATTATAAATATTGGATATTGAGTATTTTGTGA  
PIR:T04448 159

Query- 10695 AAATTTCTGTGTTTAAATCTGACTTGACTTGTGTTGTCAGTACTGACTATGCGGAACTG  
PIR:T04448 159  
T D Y A E T V

Query- 10635 TGAAATCTGCTCGTTGGGAGTATAGTACTCGTCCCGTTTACGGTTGGGGTGATGTTGGGG  
PIR:T04448 166  
K S A R W E Y S T R P V Y G W G D V G A

Query- 10575 CCAAACAGAAGTCAACTGCAGGCTGGCCTGCAGCTTTTCCTGTATTTGAGCCTCATTGGC  
PIR:T04448 186  
K Q K S T A G W P A A F P V F E P H W Q

Query- 10515 AGATATGCATGGCAGGAGGCCTTCCACAGGTGTGAGCTTTGCTTGATTGACTTAAAGTT  
PIR:T04448 206  
PIR:T04448 206  
Exon 10655 10486 Confidence: 96 100  
I C M A G G L S T G

Query- 10455 AATAAATAGACGGTTAAGTTTACTTGCCTAGTACTAACAGAAAATTAAGAAAGAAACCAC  
PIR:T04448 216

Query- 10395 CCTCTTTCTATCAGCAGAACTGCTATTGTAGTTCTTATTTTCTCTTGTATTGTCAGG  
PIR:T04448 216

Query- 10335 GTGGATAGAATGGGGCGGTGAAAGGTTTGAGTTTCGGGATGCACCTTCTTATTCAGAGAA  
PIR:T04448 216  
W I E W G G E R F E F R D A P S Y S E K

Query- 10275 GAATTGGGGTGGAGGCTTCCCAAGAAAATGGTTTGGGTAAAACATTTTCATCCTTTTGCT  
PIR:T04448 236  
PIR:T04448 236  
Exon 10336 10239 Confidence: 96 100  
N W G G G F P R K W F W

Figure 31C

35/40

Query- 10215 ACATTCTTGTGTCAGACTTTAGTTAGCTAGTGGACCTGTGTATACACCCACATGTAGTA  
PIR:T04448 248

Query- 10155 TACTTGTGTTGATAGCTTTATTTGTCAATGTCTCTTTACAGGTCCAGTGTAAATGTCTTTGA  
PIR:T04448 248  
V Q C N V F E

Query- 10095 AGGGGCAACTGGAGAAGTTGCTTTAACCAGGTGGCGGGTTGAGGCAATTGCCTGGATT  
PIR:T04448 255 G A T G E V A L T A G G G L R Q L P G L

Query- 10035 GACTGAGACCTATGAAAATGCTGCACTGGTATGCACTTATAAGATCTTCTTAAGCAATGA  
PIR:T04448 275 T E T Y E N A A L  
PIR:T04448 Exon 10115 10008 Confidence: 100 100

Query- 9975 CAGTGAGTATTAGAAGGCAGATAGTTTACAAAAGCTCTGGGCCCTTGTAATCTGCAGGT  
PIR:T04448 284  
V

Query- 9915 TTGTGTACACTATGATGGAAAAATGTACGAGTTTGTTCCTTGGAAATGGTGTGTAGATG  
PIR:T04448 285 C V H Y D G K M Y E F V P W N G V V R W  
GSDB:S:495- 532  
tagatg

Query- 9855 GGAAATGTCTCCCTGGGG TTATTGGTATATACTGCAGAGAACGAAAACCATGTGGTAA  
PIR:T04448 305 E M S P W G Y W Y I T A E N E N H V  
GSDB:S:495- 526 ggaaat tctccctgggggttattggtatataactgcagagancgNaaaccatgtg  
PIR:T04448 Exon 9917 9801 Confidence: 100 100  
GSDB:S:495- Exon 9861 9801 Confidence: 93 93

Query- 9796 ATTTGTTTTACTAGTTTTCATTGCTTTTACTTTTGACATCATATCATTCCCTTATGGCTA  
PIR:T04448 323  
GSDB:S:495- 471

Query- 9736 GATTCCAACACCCGATGAATGTCTTGTGACAGGTGGAAGTAGAGGCAAGAACAATGAAG  
PIR:T04448 323  
V E L E A R T N E A  
GSDB:S:495- 471  
gtggaactagaggcNagaacaaatgaag

Query- 9676 CGGGTACACCTCTGCGTGCTCCTACCACAGAAGTTGGGCTAGCTACGGCTTGCAGAGATA  
PIR:T04448 333 G T P L R A P T T E V G L A T A C R D S  
GSDB:S:495- 443 cgggtacacctctgctgctcctaccacagaagttgggctagctacggcttgcagagata

Query- 9616 GTTGTACGGTGAATTGAAGTTGCAGATATGGGAACGGCTATATGATGGAAGTAAAGGCA  
PIR:T04448 353 C Y G E L K L Q I W E R L Y D G S K G K  
GSDB:S:495- 383 gttgttacgggtgaattgaagttgcagatatgggaacggctatatgatggaagtaaaggca

Figure 31D

36/40

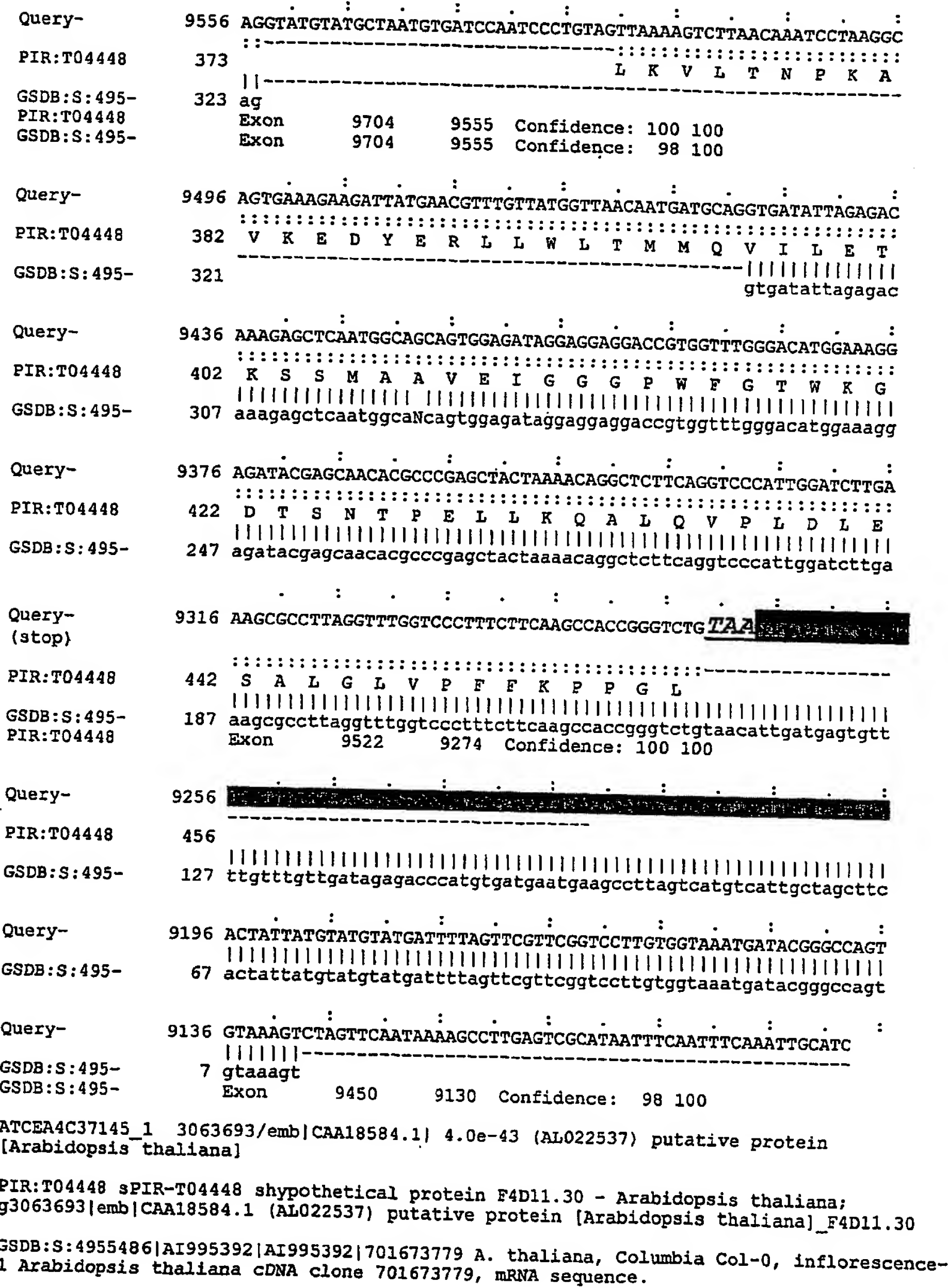


Figure 31E

37/40

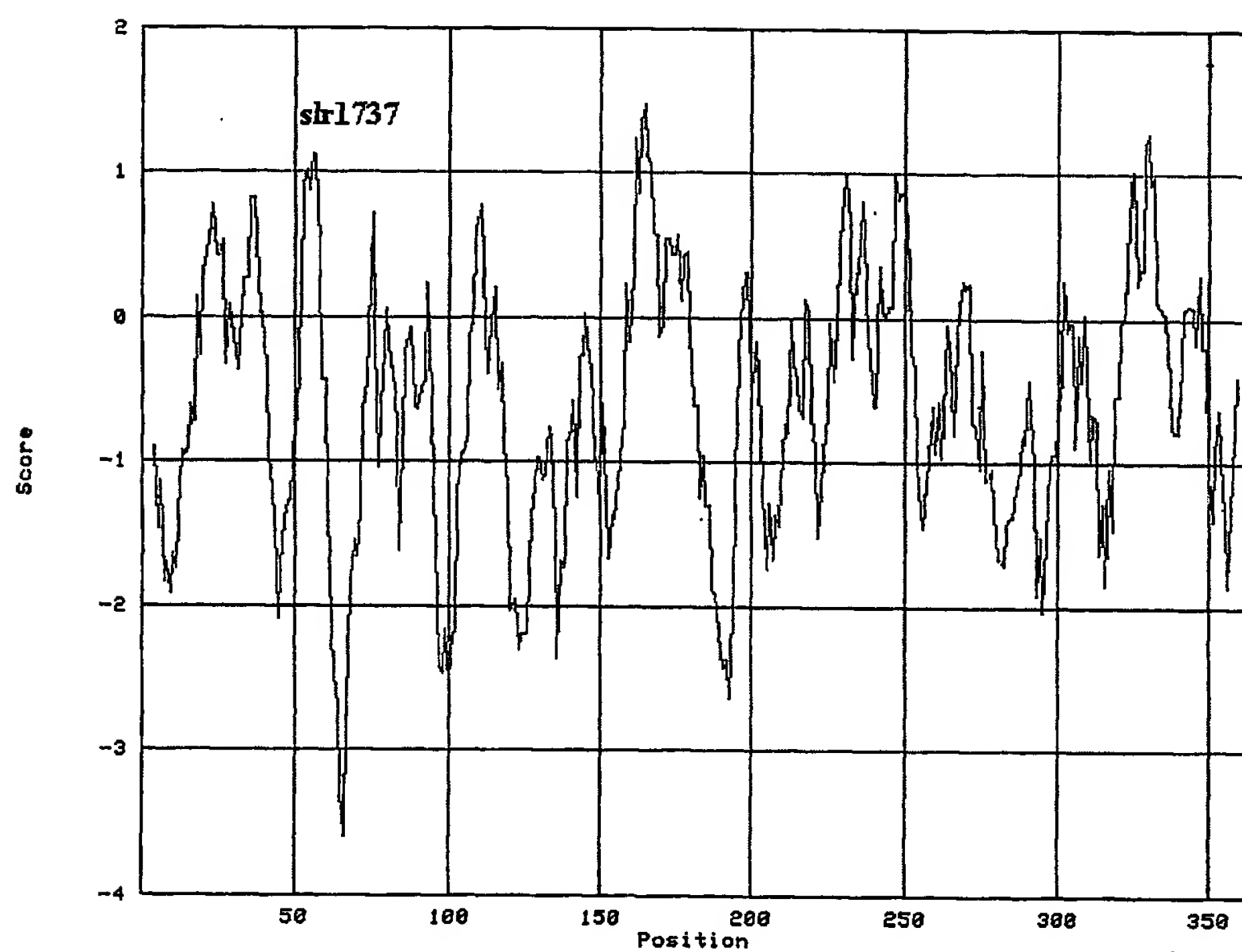


Figure 32

38/40

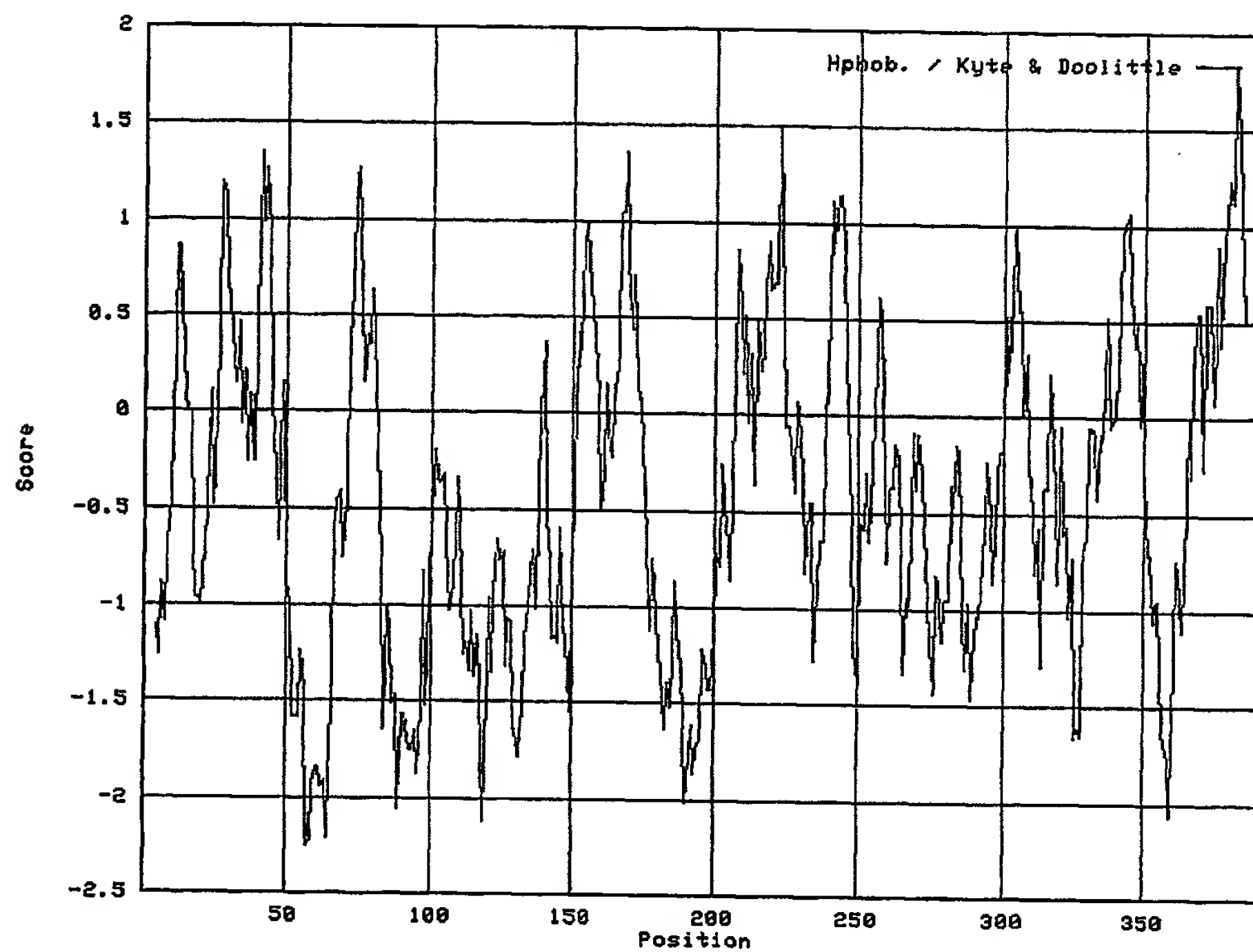


Figure 33

39/40

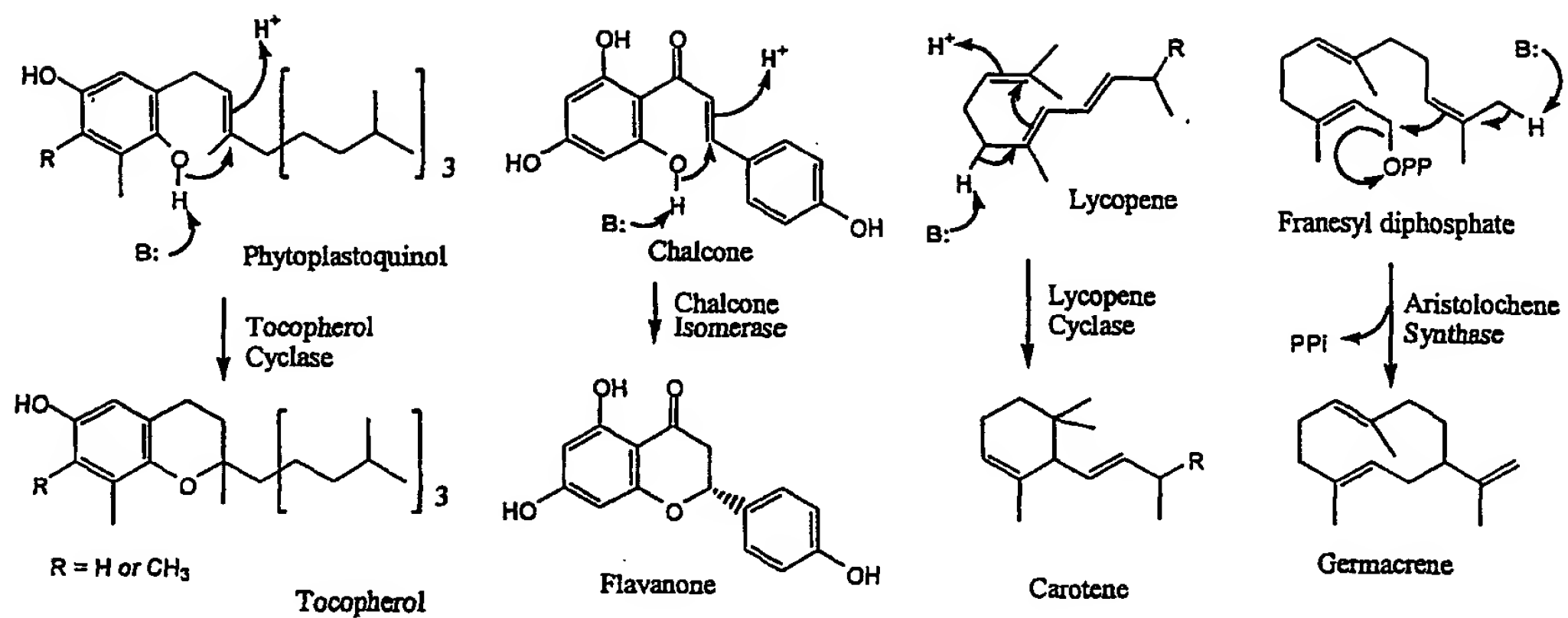


Figure 34

40/40

slr1737_SYNSP_S74814_	-----M
slr1737_ARATH_T04448_	MEIRSLIVSMNPNLSSFELSRPVSPLTRSLVPFRSTKLVPRSISRVSASI
CFI_ARATH_P41088_	-----
slr1737_SYNSP_S74814_	KFP-----PHSGYHWQGS-PFFEGWYVRL
slr1737_ARATH_T04448_	STPNSETDKISVKPVYVPTSPNRELRTPHSGYHFDGTPRKFFEGWYFRVS
CFI_ARATH_P41088_	-----
slr1737_SYNSP_S74814_	LPOSGESFAFMYSIENPASDHHYGGGAVQILGPATK----KQENQEDQLV
slr1737_ARATH_T04448_	IPEKRESFCFMYSVENPAFRQSLSPLEVALYGPRFTGVGAQILGANDKYL
CFI_ARATH_P41088_	MSSSNACASPSPPFA----VTKLHVDSV-
slr1737_SYNSP_S74814_	WRTFPSVKKFWASPRQFALG-HWGKCRDNRO-AKPLLSEEFFATVKEGYQ
slr1737_ARATH_T04448_	CQYEQDSHNFWDGRHELVLGNTFSAVPGAKAPNKEVPPEEFNRRVSEGFQ
CFI_ARATH_P41088_	--TFVPSVKSPASSNPLFLG-GAGVRGLDIQ-GK-----FVIETVIGVY
slr1737_SYNSP_S74814_	IHQNHQGGQIIHGDR-----HCRWQFTVEPEVTWGSNRFPRATAGW
slr1737_ARATH_T04448_	ATPFWHQGHICDDGRDIAETVKSARWEYSTRPVYWGWDVGAKQKSTAGW
CFI_ARATH_P41088_	LEGNAPVPSLSV-----KWKGKTTEELTESIPFFREIVTGAF
slr1737_SYNSP_S74814_	LSFLPLFDPGWQILLAQGRAHWLKWQREQYEFDHVLYAEKNWGHFSFPS
slr1737_ARATH_T04448_	PAAFPVFEPHWQICMAGGLSTGWIEWGGERFEFRDAPSYSEKNWGGGFPR
CFI_ARATH_P41088_	EKFIKVT-----M-----KLPLTGQYSEKVTENC
slr1737_SYNSP_S74814_	RWFWLQANYFPDHPG-LSVTAAGGERIVLGRPE---EVALIGLHHQGNFY
slr1737_ARATH_T04448_	KWFWVQCNVFEAGTGEVALTAGGGLRQLPGLTETYENAALVCVHYDGKMY
CFI_ARATH_P41088_	VAIWQLGLYTDCEA-KAV-----EKFLEIFKE---ET-----
slr1737_SYNSP_S74814_	EFGPGHGTVTWQVAPWGRWQLKASNDRYWVKLSGKTDKKGSLVHTP-TAQ
slr1737_ARATH_T04448_	EFVPWNGVVRWEMSPWGYWYITAENENHVVELEARTNEAGTPLRAPTTEV
CFI_ARATH_P41088_	-FPPG-SSILFALSPTGSLTVAFSKDDS-IPETGLAVIENKLLAEA-VLE
slr1737_SYNSP_S74814_	GLQLNCRDTRGYLYLQLGSGVGHG-----LIVQGETDTAGLEVGG-----
slr1737_ARATH_T04448_	GLATACRDSCYGEKLIWERLYDGSKGKVIETKSSMAAVEIGGGPWFG
CFI_ARATH_P41088_	--SIIGKNGVSPGTRLSVAERLSQ-----LMMKNKDEKEVSDHSL-----
slr1737_SYNSP_S74814_	----DWGLTEENLSKKT-----VPF-----
slr1737_ARATH_T04448_	TWKGDTSNTPELLKQALQVPLDLESALGLVPFFKPPGL
CFI_ARATH_P41088_	----EEKLAKEN-----

Figure 35



## SEQUENCE LISTING

<110> Subramaniam, Sai  
 5 Slater, Steven  
 Karberg, Katherine  
 Chen, Ridong  
 Valentin, Henry  
 Huang Wong, Yun-Hua  
 10  
 <120> Nucleic Acid Sequences Involved in  
 Tocopherol Synthesis  
 <130> MOCO.008.00WO  
 15  
 <150> 09/549,848  
 <151> 2000-04-15  
 <150> 09/688,069  
 20 <151> 2000-10-15  
 <160> 94  
 <170> FastSEQ for Windows Version 4.0  
 25  
 <210> 1  
 <211> 1182  
 <212> DNA  
 <213> Arabidopsis sp  
 30  
 <400> 1  
 atggagtctc tgctctctag ttcttctctt gtttccgctg ctggtgggtt ttggtggaag 60  
 aagcagaatc taaagctcca ctctttatca gaaatccgag ttctgcgttg tgattcgagt 120  
 aaagtgtgctg caaaaccgaa gtttaggaac aatcttggtta ggctgatgg tcaaggatct 180  
 35 tcattgttgt tgtatccaaa acataagtcg agatttcggg ttaatgccac tgcgggtcag 240  
 cctgaggctt tcgactcgaa tagcaaacag aagtctttta gagactcgtt agatgcgttt 300  
 tacaggtttt ctaggcctca tacagttatt ggcacagtgc ttagcatttt atctgtatct 360  
 ttcttagcag tagagaaggt ttctgatata tctcctttac ttttcaactgg catcttgag 420

gctgttggtg cagctctcat gatgaacatt tacatagttg ggctaaatca gttgtctgat 480  
 gttgaaatag ataagggttaa caagccctat cttccattgg catcaggaga atattctggt 540  
 aacaccggca ttgcaatagt agcttccttc tccatcatga gtttctggct tgggtggatt 600  
 gttgggttcat ggccattggt ctgggctctt tttgtgagtt tcatgctcgg tactgcatac 660  
 5 tctatcaatt tgccactttt acggtggaaa agatttgcatt tgggtgcagc aatgtgtatc 720  
 ctgctgtcc gagctattat tgttcaaata gccttttatc tacatattca gacacatgtg 780  
 tttggaagac caatcttggt cactaggcct cttattttcg ccactgcgtt tatgagcttt 840  
 ttctctgtcg ttattgcatt gtttaaggat atacctgata tcgaagggga taagatatc 900  
 ggaatccgat cattctctgt aactctgggt cagaaacggg tgttttggac atgtgttaca 960  
 10 ctacttcaaa tggcttacgc tgttgcaatt ctagttggag ccacatctcc attcatatgg 1020  
 agcaaagtca tctcggttgt gggctcatgtt atactcgcaa caactttgtg ggctcgagct 1080  
 aagtccgttg atctgagtag caaaaccgaa ataacttcat gttatatgtt catatggaag 1140  
 ctcttttatg cagagtactt gctgttacct tttttgaagt ga 1182

15 <210> 2

<211> 393

<212> PRT

<213> Arabidopsis sp

20 <400> 2

Met Glu Ser Leu Leu Ser Ser Ser Ser Leu Val Ser Ala Ala Gly Gly  
 1 5 10 15  
 Phe Cys Trp Lys Lys Gln Asn Leu Lys Leu His Ser Leu Ser Glu Ile  
 20 25 30  
 25 Arg Val Leu Arg Cys Asp Ser Ser Lys Val Val Ala Lys Pro Lys Phe  
 35 40 45  
 Arg Asn Asn Leu Val Arg Pro Asp Gly Gln Gly Ser Ser Leu Leu Leu  
 50 55 60  
 Tyr Pro Lys His Lys Ser Arg Phe Arg Val Asn Ala Thr Ala Gly Gln  
 30 65 70 75 80  
 Pro Glu Ala Phe Asp Ser Asn Ser Lys Gln Lys Ser Phe Arg Asp Ser  
 85 90 95  
 Leu Asp Ala Phe Tyr Arg Phe Ser Arg Pro His Thr Val Ile Gly Thr  
 100 105 110  
 35 Val Leu Ser Ile Leu Ser Val Ser Phe Leu Ala Val Glu Lys Val Ser  
 115 120 125  
 Asp Ile Ser Pro Leu Leu Phe Thr Gly Ile Leu Glu Ala Val Val Ala  
 130 135 140  
 Ala Leu Met Met Asn Ile Tyr Ile Val Gly Leu Asn Gln Leu Ser Asp

145                                      150                                      155                                      160  
 Val Glu Ile Asp Lys Val Asn Lys Pro Tyr Leu Pro Leu Ala Ser Gly  
    165                                      170                                      175  
 Glu Tyr Ser Val Asn Thr Gly Ile Ala Ile Val Ala Ser Phe Ser Ile  
 5                                      180                                      185                                      190  
 Met Ser Phe Trp Leu Gly Trp Ile Val Gly Ser Trp Pro Leu Phe Trp  
    195                                      200                                      205  
 Ala Leu Phe Val Ser Phe Met Leu Gly Thr Ala Tyr Ser Ile Asn Leu  
    210                                      215                                      220  
 10 Pro Leu Leu Arg Trp Lys Arg Phe Ala Leu Val Ala Ala Met Cys Ile  
    225                                      230                                      235                                      240  
 Leu Ala Val Arg Ala Ile Ile Val Gln Ile Ala Phe Tyr Leu His Ile  
    245                                      250                                      255  
 Gln Thr His Val Phe Gly Arg Pro Ile Leu Phe Thr Arg Pro Leu Ile  
 15                                      260                                      265                                      270  
 Phe Ala Thr Ala Phe Met Ser Phe Phe Ser Val Val Ile Ala Leu Phe  
    275                                      280                                      285  
 Lys Asp Ile Pro Asp Ile Glu Gly Asp Lys Ile Phe Gly Ile Arg Ser  
    290                                      295                                      300  
 20 Phe Ser Val Thr Leu Gly Gln Lys Arg Val Phe Trp Thr Cys Val Thr  
    305                                      310                                      315                                      320  
 Leu Leu Gln Met Ala Tyr Ala Val Ala Ile Leu Val Gly Ala Thr Ser  
    325                                      330                                      335  
 Pro Phe Ile Trp Ser Lys Val Ile Ser Val Val Gly His Val Ile Leu  
 25                                      340                                      345                                      350  
 Ala Thr Thr Leu Trp Ala Arg Ala Lys Ser Val Asp Leu Ser Ser Lys  
    355                                      360                                      365  
 Thr Glu Ile Thr Ser Cys Tyr Met Phe Ile Trp Lys Leu Phe Tyr Ala  
    370                                      375                                      380  
 30 Glu Tyr Leu Leu Leu Pro Phe Leu Lys  
    385                                      390

&lt;210&gt; 3

&lt;211&gt; 1224

35 &lt;212&gt; DNA

&lt;213&gt; Arabidopsis sp

&lt;400&gt; 3

atggcggtttt ttgggctctc ccgtgtttca agacgggttg tgaatcttc cgtctccgta

60

actccatctt cttcctctgc tcttttgcaa tcacaacata aatccttgtc caatcctgtg 120  
 actacccatt acacaaatcc tttcactaag tggtatcctt catggaatga taattaccaa 180  
 gtatggagta aaggaagaga attgcatcag gagaagtttt ttggtggttg ttggaattac 240  
 agattaattt gtggaatgtc gtcgtcttct tcggttttgg agggaaagcc gaagaaagat 300  
 5 gataaggaga agagtgatgg tgttggtgtt aagaaagctt cttggataga tttgtattta 360  
 ccagaagaag ttagagggtta tgctaagctt gctcgattgg ataaacccat tggaacttgg 420  
 ttgcttgctg ggccttgat gtggtcgatt gcgttggtg ctgacctgg aagccttcca 480  
 agttttaaat atatggcttt atttggttgc ggagcattac ttcttagagg tgctggttgt 540  
 actataaatg atctgcttga tcaggacata gatacaaagg ttgatcgta aaaactaaga 600  
 10 cctatcgcca gtggtctttt gacaccattt caagggattg gatttctcgg gctgcagttg 660  
 cttttaggct tagggattct tctccaactt aacaattaca gccgtgtttt aggggcttca 720  
 tctttgttac ttgtcttttc ctaccactt atgaagaggt ttacattttg gcctcaagcc 780  
 tttttagggt tgaccataaa ctggggagca ttgttaggat ggactgcagt taaaggaagc 840  
 atagcaccat ctattgtact ccctctctat ctctccggag tctgctggac ccttgtttat 900  
 15 gatactattt atgcacatca ggacaaagaa gatgatgtaa aagttggtgt taagtcaaca 960  
 gcccttagat tcggtgataa tacaagctt tggttaactg gatttggcac agcatccata 1020  
 ggttttcttg cactttctgg attcagtgca gatctcgggt ggcaatatta cgcactcactg 1080  
 gccgctgcat caggacagtt aggatggcaa atagggacag ctgacttatc atctggtgct 1140  
 gactgcagta gaaaatttgt gtcgaacaag tggtttggtg ctattatatt tagtggagtt 1200  
 20 gtacttggaa gaagttttca ataa 1224

&lt;210&gt; 4

&lt;211&gt; 407

&lt;212&gt; PRT

25 &lt;213&gt; Arabidopsis sp

&lt;400&gt; 4

Met Ala Phe Phe Gly Leu Ser Arg Val Ser Arg Arg Leu Leu Lys Ser  
 1 5 10 15  
 30 Ser Val Ser Val Thr Pro Ser Ser Ser Ser Ala Leu Leu Gln Ser Gln  
 20 25 30  
 His Lys Ser Leu Ser Asn Pro Val Thr Thr His Tyr Thr Asn Pro Phe  
 35 40 45  
 Thr Lys Cys Tyr Pro Ser Trp Asn Asp Asn Tyr Gln Val Trp Ser Lys  
 35 50 55 60  
 Gly Arg Glu Leu His Gln Glu Lys Phe Phe Gly Val Gly Trp Asn Tyr  
 65 70 75 80  
 Arg Leu Ile Cys Gly Met Ser Ser Ser Ser Ser Val Leu Glu Gly Lys  
 85 90 95

Pro Lys Lys Asp Asp Lys Glu Lys Ser Asp Gly Val Val Val Lys Lys  
 100 105 110  
 Ala Ser Trp Ile Asp Leu Tyr Leu Pro Glu Glu Val Arg Gly Tyr Ala  
 115 120 125  
 5 Lys Leu Ala Arg Leu Asp Lys Pro Ile Gly Thr Trp Leu Leu Ala Trp  
 130 135 140  
 Pro Cys Met Trp Ser Ile Ala Leu Ala Ala Asp Pro Gly Ser Leu Pro  
 145 150 155 160  
 Ser Phe Lys Tyr Met Ala Leu Phe Gly Cys Gly Ala Leu Leu Leu Arg  
 10 165 170 175  
 Gly Ala Gly Cys Thr Ile Asn Asp Leu Leu Asp Gln Asp Ile Asp Thr  
 180 185 190  
 Lys Val Asp Arg Thr Lys Leu Arg Pro Ile Ala Ser Gly Leu Leu Thr  
 195 200 205  
 15 Pro Phe Gln Gly Ile Gly Phe Leu Gly Leu Gln Leu Leu Leu Gly Leu  
 210 215 220  
 Gly Ile Leu Leu Gln Leu Asn Asn Tyr Ser Arg Val Leu Gly Ala Ser  
 225 230 235 240  
 Ser Leu Leu Leu Val Phe Ser Tyr Pro Leu Met Lys Arg Phe Thr Phe  
 20 245 250 255  
 Trp Pro Gln Ala Phe Leu Gly Leu Thr Ile Asn Trp Gly Ala Leu Leu  
 260 265 270  
 Gly Trp Thr Ala Val Lys Gly Ser Ile Ala Pro Ser Ile Val Leu Pro  
 275 280 285  
 25 Leu Tyr Leu Ser Gly Val Cys Trp Thr Leu Val Tyr Asp Thr Ile Tyr  
 290 295 300  
 Ala His Gln Asp Lys Glu Asp Asp Val Lys Val Gly Val Lys Ser Thr  
 305 310 315 320  
 Ala Leu Arg Phe Gly Asp Asn Thr Lys Leu Trp Leu Thr Gly Phe Gly  
 30 325 330 335  
 Thr Ala Ser Ile Gly Phe Leu Ala Leu Ser Gly Phe Ser Ala Asp Leu  
 340 345 350  
 Gly Trp Gln Tyr Tyr Ala Ser Leu Ala Ala Ala Ser Gly Gln Leu Gly  
 355 360 365  
 35 Trp Gln Ile Gly Thr Ala Asp Leu Ser Ser Gly Ala Asp Cys Ser Arg  
 370 375 380  
 Lys Phe Val Ser Asn Lys Trp Phe Gly Ala Ile Ile Phe Ser Gly Val  
 385 390 395 400  
 Val Leu Gly Arg Ser Phe Gln

405

<210> 5  
 <211> 1296  
 5 <212> DNA  
 <213> Arabidopsis sp

<400> 5

10	atgtggcgaa gatctgttgt ttctcgttta tcttcaagaa tctctgtttc ttcttcgtta	60
	ccaaacccta gactgattcc ttggccccgc gaattatgtg ccgttaatag cttctcccag	120
	cctccggtct cgacggaatc aactgctaag ttagggatca ctggtgtag atctgatgcc	180
	aatecgagttt ttgccactgc tactgccgcc gctacagcta cagctaccac cggtagagatt	240
	tcgtctagag ttgcggcttt ggctggatta gggcatcact acgctcgttg ttattgggag	300
	ctttctaaag ctaaacttag tatgcttgtg gttgcaactt ctggaactgg gtatattctg	360
15	ggtacgggaa atgctgcaat tagcttcccc gggctttgtt acacatgtgc aggaaccatg	420
	atgattgctg catctgctaa ttccctgaat cagatTTTTg agataagcaa tgattctaag	480
	atgaaaagaa cgatgctaag gccattgcct tcaggacgta ttagtggtcc acacgctgtt	540
	gcatgggcta ctattgctgg tgcttctggg gcttgtttgt tggccagcaa gactaatatg	600
	ttggctgctg gacttgcac tgccaatctt gtactttatg cgtttgttta tactccgttg	660
20	aagcaacttc accctatcaa tacatgggtt ggcgctgttg ttgggtgctat cccacccttg	720
	cttgggtggg cggcagcgtc tggtcagatt tcatacaatt cgatgattct tccagctgct	780
	ctttactttt ggcagatacc tcattttatg gcccttgcaac atctctgccg caatgattat	840
	gcagctggag gttacaagat gttgtcactc tttgatccgt cagggagag aatagcagca	900
	gtggctctaa ggaactgctt ttacatgac cctctcgggt tcatcgcccta tgactggggg	960
25	ttaacctcaa gttgggttttg cctcgaatca acactttctca cactagcaat cgctgcaaca	1020
	gcattttcat tctaccgaga ccggaccatg cataaagcaa ggaaaatgtt ccatgccagt	1080
	cttctcttcc ttccctgttt catgtctggg cttcttctac accgtgtctc taatgataat	1140
	cagcaacaac tcgtagaaga agccggatta acaaattctg tatctgggtga agtcaaaaact	1200
	cagaggcgaa agaaacgtgt ggctcaacct ccgggtggctt atgcctctgc tgcaccgttt	1260
30	cctttcctcc cagctccttc cttctactct ccatga	1296

<210> 6  
 <211> 431  
 <212> PRT  
 35 <213> Arabidopsis sp

<400> 6

Met	Trp	Arg	Arg	Ser	Val	Val	Tyr	Arg	Phe	Ser	Ser	Arg	Ile	Ser	Val
1					5										
															15

Ser Ser Ser Leu Pro Asn Pro Arg Leu Ile Pro Trp Ser Arg Glu Leu  
 20 25 30  
 Cys Ala Val Asn Ser Phe Ser Gln Pro Pro Val Ser Thr Glu Ser Thr  
 35 40 45  
 5 Ala Lys Leu Gly Ile Thr Gly Val Arg Ser Asp Ala Asn Arg Val Phe  
 50 55 60  
 Ala Thr Ala Thr Ala Ala Ala Thr Ala Thr Ala Thr Thr Gly Glu Ile  
 65 70 75 80  
 Ser Ser Arg Val Ala Ala Leu Ala Gly Leu Gly His His Tyr Ala Arg  
 10 85 90 95  
 Cys Tyr Trp Glu Leu Ser Lys Ala Lys Leu Ser Met Leu Val Val Ala  
 100 105 110  
 Thr Ser Gly Thr Gly Tyr Ile Leu Gly Thr Gly Asn Ala Ala Ile Ser  
 115 120 125  
 15 Phe Pro Gly Leu Cys Tyr Thr Cys Ala Gly Thr Met Met Ile Ala Ala  
 130 135 140  
 Ser Ala Asn Ser Leu Asn Gln Ile Phe Glu Ile Ser Asn Asp Ser Lys  
 145 150 155 160  
 Met Lys Arg Thr Met Leu Arg Pro Leu Pro Ser Gly Arg Ile Ser Val  
 165 170 175  
 20 Pro His Ala Val Ala Trp Ala Thr Ile Ala Gly Ala Ser Gly Ala Cys  
 180 185 190  
 Leu Leu Ala Ser Lys Thr Asn Met Leu Ala Ala Gly Leu Ala Ser Ala  
 195 200 205  
 25 Asn Leu Val Leu Tyr Ala Phe Val Tyr Thr Pro Leu Lys Gln Leu His  
 210 215 220  
 Pro Ile Asn Thr Trp Val Gly Ala Val Val Gly Ala Ile Pro Pro Leu  
 225 230 235 240  
 Leu Gly Trp Ala Ala Ala Ser Gly Gln Ile Ser Tyr Asn Ser Met Ile  
 245 250 255  
 30 Leu Pro Ala Ala Leu Tyr Phe Trp Gln Ile Pro His Phe Met Ala Leu  
 260 265 270  
 Ala His Leu Cys Arg Asn Asp Tyr Ala Ala Gly Gly Tyr Lys Met Leu  
 275 280 285  
 35 Ser Leu Phe Asp Pro Ser Gly Lys Arg Ile Ala Ala Val Ala Leu Arg  
 290 295 300  
 Asn Cys Phe Tyr Met Ile Pro Leu Gly Phe Ile Ala Tyr Asp Trp Gly  
 305 310 315 320  
 Leu Thr Ser Ser Trp Phe Cys Leu Glu Ser Thr Leu Leu Thr Leu Ala



325 330 335  
 Ile Ala Ala Thr Ala Phe Ser Phe Tyr Arg Asp Arg Thr Met His Lys  
 340 345 350  
 Ala Arg Lys Met Phe His Ala Ser Leu Leu Phe Leu Pro Val Phe Met  
 5 355 360 365  
 Ser Gly Leu Leu Leu His Arg Val Ser Asn Asp Asn Gln Gln Gln Leu  
 370 375 380  
 Val Glu Glu Ala Gly Leu Thr Asn Ser Val Ser Gly Glu Val Lys Thr  
 385 390 395 400  
 10 Gln Arg Arg Lys Lys Arg Val Ala Gln Pro Pro Val Ala Tyr Ala Ser  
 405 410 415  
 Ala Ala Pro Phe Pro Phe Leu Pro Ala Pro Ser Phe Tyr Ser Pro  
 420 425 430

15 <210> 7  
 <211> 479  
 <212> DNA  
 <213> Arabidopsis sp

20 <400> 7  
 ggaaactccc ggagcacctg tttgcaggta ccgctaacct taatcgataa tttattttctc 60  
 ttgtcaggaa ttatgtaagt ctggtggaag gctcgcatac cattttttgca ttgccttttcg 120  
 ctatgatcgg gtttactttg ggtgtgatga gaccaggcgt ggcttttatgg tatggcgaaa 180  
 acccattttt atccaatgct gcattccctc ccgatgattc gttctttcat tcctatacag 240  
 25 gtatcatgct gataaaactg ttactggtac tggtttgtat ggtatcagca agaagcgcg 300  
 cgatggcggt taaccggtat ctgcacaggc attttgacgc gaagaaccgc cgtactgcca 360  
 tccgtgaaat acctgcgggc gtcatatctg ccaacagtgc gctggtgttt acgataggct 420  
 gctgcgtggt attctgggtg gcctgttatt tcattaacac gatctgtttt tacctggcg 479

30 <210> 8  
 <211> 551  
 <212> DNA  
 <213> Arabidopsis sp

35 <220>  
 <221> misc\_feature  
 <222> (1) ... (551)  
 <223> n = A,T,C or G

&lt;400&gt; 8

ttgtggctta caccttaatg agcatagcc agnccattac ggctcggttaa tcggcgccat 60  
 ngccggngct gntgcaccgg tagtgggcta ctgcgccgtg accaatcagc ttgatctagc 120  
 ggctcttatt ctgttttttaa ttttactgtt ctggcaaagt cgcgattttt acgcgatttc 180  
 5 ctttttcagg ctaaaagact tttcagcggc ctgtattccg gtgctgcca tcattaaaga 240  
 cctgcgctat accaaaatca gcatgctggg ttacgtgggc ttatttacac tggctgctat 300  
 catgccggcc ctcttagggg atgccggtg gatttatggg atagcggcct taattttagg 360  
 cttgtattgg ctttatattg ccatacaagg attcaagacc gccgatgac aaaaatgggc 420  
 tcgtaagatg tttggatctt cgattttaat cattaccctc ttgtcggtaa tgatgcttgt 480  
 10 ttaaaacttac tgcctcctga agtttatata tcgataattt cagcttaagg aggcttagtg 540  
 gttaattcaa t 551

&lt;210&gt; 9

&lt;211&gt; 297

15 &lt;212&gt; PRT

&lt;213&gt; Arabidopsis sp

&lt;400&gt; 9

Met Val Leu Ala Glu Val Pro Lys Leu Ala Ser Ala Ala Glu Tyr Phe  
 20 1 5 10 15  
 Phe Lys Arg Gly Val Gln Gly Lys Gln Phe Arg Ser Thr Ile Leu Leu  
 20 25 30  
 Leu Met Ala Thr Ala Leu Asn Val Arg Val Pro Glu Ala Leu Ile Gly  
 35 40 45  
 25 Glu Ser Thr Asp Ile Val Thr Ser Glu Leu Arg Val Arg Gln Arg Gly  
 50 55 60  
 Ile Ala Glu Ile Thr Glu Met Ile His Val Ala Ser Leu Leu His Asp  
 65 70 75 80  
 Asp Val Leu Asp Asp Ala Asp Thr Arg Arg Gly Val Gly Ser Leu Asn  
 30 85 90 95  
 Val Val Met Gly Asn Lys Val Val Ala Leu Leu Ala Thr Ala Val Glu  
 100 105 110  
 His Leu Val Thr Gly Glu Thr Met Glu Ile Thr Ser Ser Thr Glu Gln  
 115 120 125  
 35 Arg Tyr Ser Met Asp Tyr Tyr Met Gln Lys Thr Tyr Tyr Lys Thr Ala  
 130 135 140  
 Ser Leu Ile Ser Asn Ser Cys Lys Ala Val Ala Val Leu Thr Gly Gln  
 145 150 155 160  
 Thr Ala Glu Val Ala Val Leu Ala Phe Glu Tyr Gly Arg Asn Leu Gly

165 170 175  
 Leu Ala Phe Gln Leu Ile Asp Asp Ile Leu Asp Phe Thr Gly Thr Ser  
 180 185 190  
 Ala Ser Leu Gly Lys Gly Ser Leu Ser Asp Ile Arg His Gly Val Ile  
 5 195 200 205  
 Thr Ala Pro Ile Leu Phe Ala Met Glu Glu Phe Pro Gln Leu Arg Glu  
 210 215 220  
 Val Val Asp Gln Val Glu Lys Asp Pro Arg Asn Val Asp Ile Ala Leu  
 225 230 235 240  
 10 Glu Tyr Leu Gly Lys Ser Lys Gly Ile Gln Arg Ala Arg Glu Leu Ala  
 245 250 255  
 Met Glu His Ala Asn Leu Ala Ala Ala Ala Ile Gly Ser Leu Pro Glu  
 260 265 270  
 Thr Asp Asn Glu Asp Val Lys Arg Ser Arg Arg Ala Leu Ile Asp Leu  
 15 275 280 285  
 Thr His Arg Val Ile Thr Arg Asn Lys  
 290 295

<210> 10

20 <211> 561

<212> DNA

<213> Arabidopsis sp

<400> 10

25 aagcgcaccc gtcctcttct acgattgccg ccagccgcat gtatggctgc ataaccgacc 60  
 gccctatcc gctcgcggcc gcggtcgaat tcattcacac cgcgacgctg ctgcatgacg 120  
 acgtcgtcga tgaaagcgat ttgcgccgcg gccgcgaaag cgcgcataag gttttcggca 180  
 atcaggcgag cgtgctcgtc ggcgatttcc ttttctcccg cgccttccag ctgatggtgg 240  
 aagacggctc gctcgacgcg ctgcgcattc tctcggatgc ctccgccgtg atcgcgcagg 300  
 30 gcgaagtgat gcagctcggc accgcgcgca atcttgaaac caatatgagc cagtatctcg 360  
 atgtgatcag cgcgaagacc gccgcgctct ttgccgccgc ctgcgaaatc ggcccgggtga 420  
 tggcgaacgc gaaggcggaa gatgctgccg cgatgtgcga atacggcatg aatctcggta 480  
 tcgccttcca gatcatcgac gaccttctcg attacggcac cggcggccac gccgagcttg 540  
 gcaagaacac gggcgacgat t 561

35

<210> 11

<211> 966

<212> DNA

<213> Arabidopsis sp

<400> 11

	atggtacttg ccgaggttcc aaagcttgcc tctgctgctg agtacttctt caaaaggggt	60
	gtgcaaggaa aacagtttctg ttcaactatt ttgctgctga tggcgacagc tctgaatgta	120
5	cgcggtccag aagcattgat tggggaatca acagatatag tcacatcaga attacgcgta	180
	aggcaacggg gtattgctga aatcactgaa atgatacacg tcgcaagtct actgcacgat	240
	gatgtcttgg atgatgccga tacaaggcgt ggtggttggtt ccttaaagtgt tgtaatgggt	300
	aacaagatgt cgggtattagc aggagacttc ttgctctccc gggcttggtg ggctctcgct	360
	gctttaaaga acacagaggt tgtagcatta cttgcaactg ctgtagaaca tcttgttacc	420
10	ggtgaaacca tggaaataac tagttcaacc gagcagcgtt atagtatgga ctactacatg	480
	cagaagacat attataagac agcatcgcta atctctaaca gctgcaaagc tggtgccgtt	540
	ctcactggac aaacagcaga agttgccgtg ttagcttttg agtatgggag gaatctgggt	600
	ttagcattcc aattaataga cgacattctt gatttcacgg gcacatctgc ctctctcgga	660
	aagggatcgt tgtcagatat tcgccatgga gtcataacag ccccaatcct ctttgccatg	720
15	gaagagtttc ctcaactacg cgaagttggt gatcaagttg aaaaagatcc taggaatggt	780
	gacattgctt tagagtatct tgggaagagc aagggaatac agagggcaag agaattagcc	840
	atggaacatg cgaatctagc agcagctgca atcgggtctc tacctgaaac agacaatgaa	900
	gatgtcaaaa gatcgaggcg ggcacttatt gacttgaccc atagagtcac caccagaaac	960
	aagtga	966

20

&lt;210&gt; 12

&lt;211&gt; 321

&lt;212&gt; PRT

&lt;213&gt; Arabidopsis sp

25

&lt;400&gt; 12

	Met Val Leu Ala Glu Val Pro Lys Leu Ala Ser Ala Ala Glu Tyr Phe	
	1	5 10 15
	Phe Lys Arg Gly Val Gln Gly Lys Gln Phe Arg Ser Thr Ile Leu Leu	
30	20 25 30	
	Leu Met Ala Thr Ala Leu Asn Val Arg Val Pro Glu Ala Leu Ile Gly	
	35 40 45	
	Glu Ser Thr Asp Ile Val Thr Ser Glu Leu Arg Val Arg Gln Arg Gly	
	50 55 60	
35	Ile Ala Glu Ile Thr Glu Met Ile His Val Ala Ser Leu Leu His Asp	
	65 70 75 80	
	Asp Val Leu Asp Asp Ala Asp Thr Arg Arg Gly Val Gly Ser Leu Asn	
	85 90 95	
	Val Val Met Gly Asn Lys Met Ser Val Leu Ala Gly Asp Phe Leu Leu	

100 105 110 ,  
 Ser Arg Ala Cys Gly Ala Leu Ala Ala Leu Lys Asn Thr Glu Val Val  
 115 120 125  
 Ala Leu Leu Ala Thr Ala Val Glu His Leu Val Thr Gly Glu Thr Met  
 5 130 135 140  
 Glu Ile Thr Ser Ser Thr Glu Gln Arg Tyr Ser Met Asp Tyr Tyr Met  
 145 150 155 160  
 Gln Lys Thr Tyr Tyr Lys Thr Ala Ser Leu Ile Ser Asn Ser Cys Lys  
 165 170 175  
 10 Ala Val Ala Val Leu Thr Gly Gln Thr Ala Glu Val Ala Val Leu Ala  
 180 185 190  
 Phe Glu Tyr Gly Arg Asn Leu Gly Leu Ala Phe Gln Leu Ile Asp Asp  
 195 200 205  
 Ile Leu Asp Phe Thr Gly Thr Ser Ala Ser Leu Gly Lys Gly Ser Leu  
 15 210 215 220  
 Ser Asp Ile Arg His Gly Val Ile Thr Ala Pro Ile Leu Phe Ala Met  
 225 230 235 240  
 Glu Glu Phe Pro Gln Leu Arg Glu Val Val Asp Gln Val Glu Lys Asp  
 245 250 255  
 20 Pro Arg Asn Val Asp Ile Ala Leu Glu Tyr Leu Gly Lys Ser Lys Gly  
 260 265 270  
 Ile Gln Arg Ala Arg Glu Leu Ala Met Glu His Ala Asn Leu Ala Ala  
 275 280 285  
 Ala Ala Ile Gly Ser Leu Pro Glu Thr Asp Asn Glu Asp Val Lys Arg  
 25 290 295 300  
 Ser Arg Arg Ala Leu Ile Asp Leu Thr His Arg Val Ile Thr Arg Asn  
 305 310 315 320  
 Lys  
 30  
 <210> 13  
 <211> 621  
 <212> DNA  
 <213> Arabidopsis sp  
 35  
 <400> 13  
 gctttctcct ttgctaattc ttgagctttc ttgatccac cgcgatttct aactatttca 60  
 atcgcttctt caagcgatcc aggtcacaa aactcagact caatgatctc tottagcctt 120  
 ggctcattct ctagcgcgaa gatcactggc gccgttatgt tacctttggc taagtcatta 180

	gctgcaggct tacctaactg ctctgtggac tgagtgaagt ccagaatgtc atcaactact	240
	tgaaaagata aaccgagatt cttcccgaac tgatacattt gctctgcgac cttgctttcg	300
	actttactga aaattgctgc tcctttggtg cttgcagcta ctaatgaagc tgtctttag	360
	taactcttta gcatgtagtc atcaagcttg acatcacaat cgaataaact cgatgcttgc	420
5	tttatctcac cgcttgcaaa atctttgatc acctgcaaaa agataaatca agattcagac	480
	caaatgttct ttgtattgag tagcttcac taatctcaga aaggaatatt acctgactta	540
	tgagcttaat gacttcaagg ttttcgagat ttgtaagtac catgatgctt gagcaacatg	600
	aaatccccag ctaatacagc t	621
10	<210> 14	
	<211> 741	
	<212> DNA	
	<213> Arabidopsis sp	
15	<400> 14	
	ggtgagtttt gttaatagtt atgagattca tctatTTTTg tcataaaaatt gtttggtttg	60
	gtttaaactc tgtgtataat tgcaggaaag gaaacagttc atgagctttt cggcacaaga	120
	gtagcgggtgc tagctggaga tttcatgttt gctcaagcgt catggtactt agcaaacttc	180
	gagaatcttg aagttattaa gctcatcagt caggtactta gttactctta cattgttttt	240
20	ctatgaggtt gagctatgaa tctcatttcg ttgaataatg ctgtgcctca aactTTTTTT	300
	catgttttca ggtgatcaaa gactttgcaa gcggagagat aaagcaggcg tccagcttat	360
	ttgactgcga caccaagctc gacgagtact tactcaaaag tttctacaag acagcctctt	420
	tagtggtgc gagcaccaaa ggagctgcc a ttttcagcag agttgagcct gatgtgacag	480
	aacaaatgta cgagtttggg aagaatctcg gtctctcttt ccagatagtt gatgatattt	540
25	tggatttcac tcagtcgaca gacgagctcg ggaagccagc agggagtgat ttggctaaag	600
	gtaacttaac agcacctgtg attttcgctc tggagagggg gccaaaggcta agagagatca	660
	ttgagtcaaa gttctgtgag gcgggttctc tggaagaagc gattgaagcg gtgacaaaag	720
	gtggggggat taagagagca c	741
30	<210> 15	
	<211> 1087	
	<212> DNA	
	<213> Arabidopsis sp	
35	<400> 15	
	cctcttcagc caatccagag gaagaagaga caacttttta tctttcgtca agagtctccg	60
	aaaacgcacg gttttatgct ctctcttctg cctcacctc acaagacgca gggcacatga	120
	ttcaaccaga gggaaaaagc aacgataaca actctgcttt tgatttcaag ctgtatatga	180
	tccgcaaagc cgagtctgta aatgcggctc tcgacgtttc cgtaccgctt ctgaaacccc	240

	ttacgatcca agaagcggtc aggtactctt tgctagccgg cggaaaacgt gtgaggcctc	300
	tgctctgcat tgccgcttgt gagcttgtgg ggggcgacga ggctactgcc atgtcagccg	360
	cttgcgcggt cgagatgata cacacaagct ctctcattca tgacgatctt ccgtgcatgg	420
	acaatgccga cctccgtaga ggcaagccca ccaatcacia ggtatgttgt ttaattatat	480
5	gaaggctcag agataatgct gaactagtgt tgaaccaatt tttgctcaaa caaggtatat	540
	ggagaagaca tggcggtttt ggcaggtgat gcactccttg cattggcggt tgagcacatg	600
	acggttgtgt cgagtgggtt ggtcgctccc gagaagatga ttcgcgccgt ggttgagctg	660
	gccagggcca tagggactac agggctagtt gctggacaaa tgatagacct agccagcgaa	720
	agactgaatc cagacaaggt tggattggag catctagagt tcatccatct ccacaaaacg	780
10	gcggcattgt tggaggcagc ggcagtttta ggggttataa tgggaggtgg aacagaggaa	840
	gaaatcgaaa agcttagaaa gtatgctagg tgtattggac tactgtttca ggttggtgat	900
	gacattctcg acgtaacaaa atctactgag gaattgggta agacagccgg aaaagacgta	960
	atggccggaa agctgacgta tccaaggctg ataggtttgg agggatccag ggaagttgca	1020
	gagcacctga ggagagaagc agaggaaaag cttaaagggt ttgatccaag tcaggcgggc	1080
15	cctctgg	1087
	<210> 16	
	<211> 1164	
	<212> DNA	
20	<213> Arabidopsis sp	
	<400> 16	
	atgacttcga ttctcaacac tgtctccacc atccactctt ccagagttac ctccgtcgat	60
	cgagtcggag tcctctctct tcggaattcg gattccgttg agttcactcg ccggcggtct	120
25	ggtttctcga cgttgateta cgaatcaccg gggcgagat ttgttgtgcg tgcggcggag	180
	actgatactg ataaagttaa atctcagaca cctgacaagg caccagccgg tggttcaagc	240
	attaaccagc ttctcggtat caaaggagca tctcaagaaa ctaataaatg gaagattcgt	300
	cttcagctta caaaaccagt cacttggcct cactgggtt ggggagtcgt ctgtggtgct	360
	gctgcttcag ggaactttca ttggacccca gaggatgttg ctaagtcgat tctttgcatg	420
30	atgatgtctg gtccttgtct tactggctat acacagacaa tcaacgactg gtatgataga	480
	gatatcgacg caattaatga gccatatcgt ccaattccat ctggagcaat atcagagcca	540
	gaggttatta cacaagtctg ggtgctatta ttgggaggtc ttggtattgc tggaatatta	600
	gatgtgtggg cagggcatac cactcccact gtcttctatc ttgctttggg aggatcattg	660
	ctatcttata tatactctgc tccacctctt aagctaaaac aaaatggatg ggttggaat	720
35	tttgacttg gagcaagcta tattagtttg ccatgggtgg ctggccaagc attggttggc	780
	actcttacgc cagatgttgt tgttctaaca ctcttgtaca gcatagctgg gttaggaata	840
	gccattgtta acgacttcaa aagtgttgaa ggagatagag cattaggact tcagtctctc	900
	ccagtagctt ttggcaccga aactgcaaaa tggatatgcy ttggtgctat agacattact	960
	cagctttctg ttgccggata tctattagca tctgggaaac cttattatgc gttggcggtg	1020



gttgctttga tcattcctca gattgtgttc cagtttaaact actttctcaa ggaccctgtc 1080  
 aaatacgacg tcaagtagca ggcaagcgcg cagccattct tgggtgctcgg aatatttgta 1140  
 acggcattag catcgcaaca ctga 1164

5 <210> 17

<211> 387

<212> PRT

<213> Arabidopsis sp

10 <400> 17

Met Thr Ser Ile Leu Asn Thr Val Ser Thr Ile His Ser Ser Arg Val  
 1 5 10 15

Thr Ser Val Asp Arg Val Gly Val Leu Ser Leu Arg Asn Ser Asp Ser  
 20 25 30

15 Val Glu Phe Thr Arg Arg Arg Ser Gly Phe Ser Thr Leu Ile Tyr Glu  
 35 40 45

Ser Pro Gly Arg Arg Phe Val Val Arg Ala Ala Glu Thr Asp Thr Asp  
 50 55 60

20 Lys Val Lys Ser Gln Thr Pro Asp Lys Ala Pro Ala Gly Gly Ser Ser  
 65 70 75 80

Ile Asn Gln Leu Leu Gly Ile Lys Gly Ala Ser Gln Glu Thr Asn Lys  
 85 90 95

Trp Lys Ile Arg Leu Gln Leu Thr Lys Pro Val Thr Trp Pro Pro Leu  
 100 105 110

25 Val Trp Gly Val Val Cys Gly Ala Ala Ala Ser Gly Asn Phe His Trp  
 115 120 125

Thr Pro Glu Asp Val Ala Lys Ser Ile Leu Cys Met Met Met Ser Gly  
 130 135 140

30 Pro Cys Leu Thr Gly Tyr Thr Gln Thr Ile Asn Asp Trp Tyr Asp Arg  
 145 150 155 160

Asp Ile Asp Ala Ile Asn Glu Pro Tyr Arg Pro Ile Pro Ser Gly Ala  
 165 170 175

Ile Ser Glu Pro Glu Val Ile Thr Gln Val Trp Val Leu Leu Leu Gly  
 180 185 190

35 Gly Leu Gly Ile Ala Gly Ile Leu Asp Val Trp Ala Gly His Thr Thr  
 195 200 205

Pro Thr Val Phe Tyr Leu Ala Leu Gly Gly Ser Leu Leu Ser Tyr Ile  
 210 215 220

Tyr Ser Ala Pro Pro Leu Lys Leu Lys Gln Asn Gly Trp Val Gly Asn

225                      230                      235                      240  
 Phe Ala Leu Gly Ala Ser Tyr Ile Ser Leu Pro Trp Trp Ala Gly Gln  
                          245                      250                      255  
 Ala Leu Phe Gly Thr Leu Thr Pro Asp Val Val Val Leu Thr Leu Leu  
 5                      260                      265                      270  
 Tyr Ser Ile Ala Gly Leu Gly Ile Ala Ile Val Asn Asp Phe Lys Ser  
                          275                      280                      285  
 Val Glu Gly Asp Arg Ala Leu Gly Leu Gln Ser Leu Pro Val Ala Phe  
                          290                      295                      300  
 10 Gly Thr Glu Thr Ala Lys Trp Ile Cys Val Gly Ala Ile Asp Ile Thr  
                          305                      310                      315                      320  
 Gln Leu Ser Val Ala Gly Tyr Leu Leu Ala Ser Gly Lys Pro Tyr Tyr  
                          325                      330                      335  
 Ala Leu Ala Leu Val Ala Leu Ile Ile Pro Gln Ile Val Phe Gln Phe  
 15                      340                      345                      350  
 Lys Tyr Phe Leu Lys Asp Pro Val Lys Tyr Asp Val Lys Tyr Gln Ala  
                          355                      360                      365  
 Ser Ala Gln Pro Phe Leu Val Leu Gly Ile Phe Val Thr Ala Leu Ala  
                          370                      375                      380  
 20 Ser Gln His  
                          385  
  
 <210> 18  
 <211> 981  
 25 <212> DNA  
      <213> Arabidopsis sp  
  
 <400> 18  
 atgttggttta gtgggttcagc gatcccatta agcagcttct gctctcttcc ggagaaaccc                      60  
 30 cacactcttc ctatgaaact ctctcccgt gcaatccgat cttcatcctc atctgccccg                      120  
 gggtcggtga acttcgatct gaggacgtat tggacgactc tgatcaccga gatcaaccag                      180  
 aagctggatg aggccatacc ggtcaagcac cctgcgggga tctacgaggc tatgagatac                      240  
 tctgtactcg cacaaggcgc caagcgtgcc cctcctgtga tgtgtgtggc ggcctgcgag                      300  
 ctcttcggtg gcgatcgcct cgccgctttc cccaccgcct gtgccctaga aatggtgcac                      360  
 35 goggcttcgt tgatacacga cgacctcccc tgtatggacg acgatcctgt gcgcagagga                      420  
 aagccatcta accacactgt ctacggctct ggcattggcca ttctcgccgg tgacgccctc                      480  
 ttcccactcg ccttccagca cattgtctcc cacacgcctc ctgaccttgt tccccgagcc                      540  
 accatcctca gactcatcac tgagattgcc cgcactgtcg gctccactgg tatggctgca                      600  
 ggccagtacg tcgaccttga aggaggtccc ttctctcttt cctttgttca ggagaagaaa                      660

	ttcggagcca tgggtgaatg ctctgccgtg tgcggtggcc tattgggcgg tgccactgag	720
	gatgagctcc agagtctccg aaggtacggg agagccgtcg ggatgctgta tcaggtggtc	780
	gatgacatca ccgaggacaa gaagaagagc tatgatgggtg gagcagagaa gggaatgatg	840
	gaaatggcgg aagagctcaa ggagaaggcg aagaaggagc ttcaagtgtt tgacaacaag	900
5	tatggaggag gagacacact tgttcctctc tacaccttcg ttgactacgc tgctcatcga	960
	cattttcttc ttcccctctg a	981
	<210> 19	
	<211> 245	
10	<212> DNA	
	<213> Glycine sp	
	<400> 19	
	gcaacatctg ggactggggt tgtcttgggg agtggtagtg ctgttgatct ttcggcactt	60
15	tcttgcactt gcttgggtac catgatgggt gctgcatctg ctaactcttt gaatcaggtg	120
	tttgagatca ataatgatgc taaaatgaag agaacaagtc gcaggccact accctcagga	180
	cgcatacaca tacctcatgc agttggctgg gcatactctg ttggattagc tggtagggct	240
	ctact	245
20	<210> 20	
	<211> 253	
	<212> DNA	
	<213> Glycine sp	
25	<400> 20	
	attggctttc caagatcatt gggttttctt gttgcattca tgaccttcta ctcttgggt	60
	ttggcattgt ccaaggatat acctgacgtt gaaggagata aagagcacgg cattgattct	120
	tttgcagtac gtctaggtca gaaacgggca ttttggattt gcgtttcctt ttttgaaatg	180
	gctttcggag ttggtatcct ggccggagca tcatgctcac acttttggac taaaattttc	240
30	acgggtatgg gaa	253
	<210> 21	
	<211> 275	
	<212> DNA	
35	<213> Glycine sp	
	<400> 21	
	tgatcttcta ctctctgggt atggcattgt ccaaggatat atctgacgtt aaaggagata	60
	aagcatacgg catcgatact ttagcgatac gtttgggtca aaaatgggta ttttggattt	120

gcattatcct ttttgaaatg gcttttggag ttgccctctt ggcaggagca acatcttctt 180  
 acctttggat taaaattgtc acgggtctgg gacatgctat tcttgcttca attctcttgt 240  
 accaagccaa atctatatac ttgagcaaca aagtt 275

5 <210> 22

<211> 299

<212> DNA

<213> Glycine sp

10 <220>

<221> misc\_feature

<222> (1) ... (299)

<223> n = A,T,C or G

15 <400> 22

ccanaatang tncatcttng aaagacaatt ggcctcttca acacacaagt ctgcatgtga 60  
 agaagaggcc aattgtcttt ccaagatcac ttatngtggc tattgtaatc atgaacttct 120  
 tctttgtggg tatggcattg gcaaaggata tacctanctg ttgaaggaga taaaatatat 180  
 ggcattgata cttttgcaat acgtataggt caaaaacaag tattttggat ttgtattttc 240  
 20 ctttttgaaa ggctttcgga gtttccctag tggcaggagc aacatcttct agccttggt 299

<210> 23

<211> 767

<212> DNA

25 <213> Glycine sp

<400> 23

gtggaggctg tggttgctgc cctgtttatg aatatttata ttgttggttt gaatcaattg 60  
 tctgatgttg aaatagacaa gataaacaag ccgtatcttc cattagcatc tggggaatat 120  
 30 tcctttgaaa ctgggtgtcac tattgttgca tctttttcaa ttctgagttt ttggcttggc 180  
 tgggttgtag gttcatggcc attatttttg gccctttttg taagctttgt gctaggaact 240  
 gcttattcaa tcaatgtgcc tctgttgaga tggaagaggt ttgcagtgtc tgcagcgatg 300  
 tgcattctag ctgttcgggc agtaatagtt caacttgcat ttttccttca catgcagact 360  
 catgtgtaca agaggccacc tgtcttttca agaccattga tttttgctac tgcattcatg 420  
 35 agcttcttct ctgtagttat agcactgttt aaggatatac ctgacattga aggagataaa 480  
 gtatttggca tccaatcttt ttcagtgtgt ttaggtcaga agccggtgtt ctggacttgt 540  
 gttacccttc ttgaaatagc ttatggagtc gccctcctgg tgggagctgc atctccttgt 600  
 ctttggagca aaattttcac gggctctggga cacgctgtgc tggcttcaat tctctggttt 660  
 catgccaaat ctgtagattt gaaaagcaaa gcttcgataa catccttcta tatgtttatt 720

tggaagctat tttatgcaga atacttactc attccttttg ttagatg

767

<210> 24

<211> 255

5 <212> PRT

<213> Glycine sp

<400> 24

Val Glu Ala Val Val Ala Ala Leu Phe Met Asn Ile Tyr Ile Val Gly  
 10 1 5 10 15  
 Leu Asn Gln Leu Ser Asp Val Glu Ile Asp Lys Ile Asn Lys Pro Tyr  
 20 25 30  
 Leu Pro Leu Ala Ser Gly Glu Tyr Ser Phe Glu Thr Gly Val Thr Ile  
 35 40 45  
 15 Val Ala Ser Phe Ser Ile Leu Ser Phe Trp Leu Gly Trp Val Val Gly  
 50 55 60  
 Ser Trp Pro Leu Phe Trp Ala Leu Phe Val Ser Phe Val Leu Gly Thr  
 65 70 75 80  
 Ala Tyr Ser Ile Asn Val Pro Leu Leu Arg Trp Lys Arg Phe Ala Val  
 20 85 90 95  
 Leu Ala Ala Met Cys Ile Leu Ala Val Arg Ala Val Ile Val Gln Leu  
 100 105 110  
 Ala Phe Phe Leu His Met Gln Thr His Val Tyr Lys Arg Pro Pro Val  
 115 120 125  
 25 Phe Ser Arg Pro Leu Ile Phe Ala Thr Ala Phe Met Ser Phe Phe Ser  
 130 135 140  
 Val Val Ile Ala Leu Phe Lys Asp Ile Pro Asp Ile Glu Gly Asp Lys  
 145 150 155 160  
 Val Phe Gly Ile Gln Ser Phe Ser Val Cys Leu Gly Gln Lys Pro Val  
 30 165 170 175  
 Phe Trp Thr Cys Val Thr Leu Leu Glu Ile Ala Tyr Gly Val Ala Leu  
 180 185 190  
 Leu Val Gly Ala Ala Ser Pro Cys Leu Trp Ser Lys Ile Phe Thr Gly  
 195 200 205  
 35 Leu Gly His Ala Val Leu Ala Ser Ile Leu Trp Phe His Ala Lys Ser  
 210 215 220  
 Val Asp Leu Lys Ser Lys Ala Ser Ile Thr Ser Phe Tyr Met Phe Ile  
 225 230 235 240  
 Trp Lys Leu Phe Tyr Ala Glu Tyr Leu Leu Ile Pro Phe Val Arg

245

250

255

&lt;210&gt; 25

&lt;211&gt; 360

5 &lt;212&gt; DNA

&lt;213&gt; Zea sp

&lt;220&gt;

&lt;221&gt; misc\_feature

10 &lt;222&gt; (1) ... (360)

&lt;223&gt; n = A,T,C or G

&lt;400&gt; 25

15 ggcgtcttca cttgttctgg tcttctcgta tccctgatg aagaggttca cattttggcc 60  
 tcaggcttat cttggcctga cattcaactg gggagcttta ctaggggtggg ctgctattaa 120  
 ggaaagcata gaccctgcaa atcatccttc cattgtatac agctggtatt tggtggacgc 180  
 tgggtgatga tactatatat gcgcctcagg tggttctgta tccctacttt catattaatc 240  
 cttgatgaag tggccatttc atgttgctgc ggtggtctta tacttgcata tctccatgca 300  
 tctcaggaca aagangatga cctgaaagta ggagtccaag tccacagctt aagatttggg 360

20

&lt;210&gt; 26

&lt;211&gt; 299

&lt;212&gt; DNA

&lt;213&gt; Zea sp

25

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1) ... (299)

&lt;223&gt; n = A,T,C or G

30

&lt;400&gt; 26

gatggttgca gcatctgcaa ataccctcaa ccagggtgttt gngataaaaa atgatgctaa 60  
 aatgaaaagg acaatgcgtg cccctgcca tctggctgca ttagtcctgc acatgctgcg 120  
 atgtgggcta caagtgttgg agttgcagga acagctttgt tggcctggaa ggctaattggc 180  
 35 ttggcagctg ggcttgcagc ttctaattctt gttctgtatg catttgtgta tacgccgttg 240  
 aagcaaatac accctgttaa tacatggggtt ggggcagtcg ttggtgccat cccaccact 299

&lt;210&gt; 27

&lt;211&gt; 255

&lt;212&gt; DNA

&lt;213&gt; Zea sp

&lt;220&gt;

5 &lt;221&gt; misc\_feature

&lt;222&gt; (1)...(255)

&lt;223&gt; n = A,T,C or G

&lt;400&gt; 27

10 anacttgc atctccatgc ntctcaggac aaagangatg acctgaaagt aggtgtcaag 60  
 tccacagcat taagatttgg agatttgacc nnatactgna tcagtggctt tggcgcggca 120  
 tgcctcggca gcttagcact cagtggttac aatgctgacc ttggttgggtg tttagtgtga 180  
 tgcttgagcg aagaatggta tngtttttac ttgatattga ctccagacct gaaatcatgt 240  
 tggacagggt ggccc 255

15

&lt;210&gt; 28

&lt;211&gt; 257

&lt;212&gt; DNA

&lt;213&gt; Zea sp

20

&lt;400&gt; 28

attgaagggg ataggactct ggggcttcag tcacttcctg ttgcttttgg gatggaaact 60  
 gcaaaatgga tttgtgttgg agcaattgat atcactcaat tatctgttgc aggttaccta 120  
 ttgagcaccg gtaagctgta ttatgccctg gtgttgcttg ggctaacaat tcctcagggtg 180  
 25 ttctttcagt tccagtactt cctgaaggac cctgtgaagt atgatgtcaa atatcaggca 240  
 agcgcacaac cattctt 257

&lt;210&gt; 29

&lt;211&gt; 368

30 &lt;212&gt; DNA

&lt;213&gt; Zea sp

&lt;400&gt; 29

atccagttgc aaataataat ggcgttcttc tctgttgtaa tagcactatt caaggatata 60  
 35 cctgacatcg aaggggaccg catattcggg atccgatcct tcagcgtccg gttagggcaa 120  
 aagaaggctt tttggatctg cgttggcttg cttgagatgg cctacagcgt tgcgatactg 180  
 atgggagcta cctcttcctg tttgtggagc aaaacagcaa ccatcgctgg ccattccata 240  
 cttgccgcga tcctatggag ctgcgcgcga tcggtggact tgacgagcaa agccgcaata 300  
 acgtccttct acatgttcat ctggaagctg ttctacgcgg agtacctgct catccctctg 360

gtgcggtg

368

&lt;210&gt; 30

&lt;211&gt; 122

5 &lt;212&gt; PRT

&lt;213&gt; Zea sp

&lt;400&gt; 30

Ile Gln Leu Gln Ile Ile Met Ala Phe Phe Ser Val Val Ile Ala Leu  
 10 1 5 10 15  
 Phe Lys Asp Ile Pro Asp Ile Glu Gly Asp Arg Ile Phe Gly Ile Arg  
 20 25 30  
 Ser Phe Ser Val Arg Leu Gly Gln Lys Lys Val Phe Trp Ile Cys Val  
 35 40 45  
 15 Gly Leu Leu Glu Met Ala Tyr Ser Val Ala Ile Leu Met Gly Ala Thr  
 50 55 60  
 Ser Ser Cys Leu Trp Ser Lys Thr Ala Thr Ile Ala Gly His Ser Ile  
 65 70 75 80  
 Leu Ala Ala Ile Leu Trp Ser Cys Ala Arg Ser Val Asp Leu Thr Ser  
 20 85 90 95  
 Lys Ala Ala Ile Thr Ser Phe Tyr Met Phe Ile Trp Lys Leu Phe Tyr  
 100 105 110  
 Ala Glu Tyr Leu Leu Ile Pro Leu Val Arg  
 115 120

25

&lt;210&gt; 31

&lt;211&gt; 278

&lt;212&gt; DNA

&lt;213&gt; Zea sp

30

&lt;400&gt; 31

tattcagcac cacctctcaa gctcaagcag aatggatgga ttgggaactt cgctctgggt 60  
 gcgagttaca tcagcttgcc ctgggtgggct ggccaggcgt tatttggaac tcttacacca 120  
 gatatcattg tcttgactac tttgtacagc atagctgggc tagggattgc tattgtaaat 180  
 35 gatttcaaga gtattgaagg ggataggact ctggggcttc agtcacttcc tgttgctttt 240  
 gggatggaaa ctgcaaaatg gatttgtgtt ggagcaat 278

&lt;210&gt; 32

&lt;211&gt; 292



&lt;212&gt; PRT

&lt;213&gt; Synechocystis sp

&lt;400&gt; 32

5 Met Val Ala Gln Thr Pro Ser Ser Pro Pro Leu Trp Leu Thr Ile Ile  
 1 5 10 15  
 Tyr Leu Leu Arg Trp His Lys Pro Ala Gly Arg Leu Ile Leu Met Ile  
 20 25 30  
 Pro Ala Leu Trp Ala Val Cys Leu Ala Ala Gln Gly Leu Pro Pro Leu  
 10 35 40 45  
 Pro Leu Leu Gly Thr Ile Ala Leu Gly Thr Leu Ala Thr Ser Gly Leu  
 50 55 60  
 Gly Cys Val Val Asn Asp Leu Trp Asp Arg Asp Ile Asp Pro Gln Val  
 65 70 75 80  
 15 Glu Arg Thr Lys Gln Arg Pro Leu Ala Ala Arg Ala Leu Ser Val Gln  
 85 90 95  
 Val Gly Ile Gly Val Ala Leu Val Ala Leu Leu Cys Ala Ala Gly Leu  
 100 105 110  
 Ala Phe Tyr Leu Thr Pro Leu Ser Phe Trp Leu Cys Val Ala Ala Val  
 20 115 120 125  
 Pro Val Ile Val Ala Tyr Pro Gly Ala Lys Arg Val Phe Pro Val Pro  
 130 135 140  
 Gln Leu Val Leu Ser Ile Ala Trp Gly Phe Ala Val Leu Ile Ser Trp  
 145 150 155 160  
 25 Ser Ala Val Thr Gly Asp Leu Thr Asp Ala Thr Trp Val Leu Trp Gly  
 165 170 175  
 Ala Thr Val Phe Trp Thr Leu Gly Phe Asp Thr Val Tyr Ala Met Ala  
 180 185 190  
 Asp Arg Glu Asp Asp Arg Arg Ile Gly Val Asn Ser Ser Ala Leu Phe  
 30 195 200 205  
 Phe Gly Gln Tyr Val Gly Glu Ala Val Gly Ile Phe Phe Ala Leu Thr  
 210 215 220  
 Ile Gly Cys Leu Phe Tyr Leu Gly Met Ile Leu Met Leu Asn Pro Leu  
 225 230 235 240  
 35 Tyr Trp Leu Ser Leu Ala Ile Ala Ile Val Gly Trp Val Ile Gln Tyr  
 245 250 255  
 Ile Gln Leu Ser Ala Pro Thr Pro Glu Pro Lys Leu Tyr Gly Gln Ile  
 260 265 270  
 Phe Gly Gln Asn Val Ile Ile Gly Phe Val Leu Leu Ala Gly Met Leu

275  
 Leu Gly Trp Leu  
 290

280  
 285

5 <210> 33  
 <211> 316  
 <212> PRT  
 <213> Synechocystis sp

10 <400> 33  
 Met Val Thr Ser Thr Lys Ile His Arg Gln His Asp Ser Met Gly Ala  
 1 5 10 15  
 Val Cys Lys Ser Tyr Tyr Gln Leu Thr Lys Pro Arg Ile Ile Pro Leu  
 20 25 30  
 15 Leu Leu Ile Thr Thr Ala Ala Ser Met Trp Ile Ala Ser Glu Gly Arg  
 35 40 45  
 Val Asp Leu Pro Lys Leu Leu Ile Thr Leu Leu Gly Gly Thr Leu Ala  
 50 55 60  
 Ala Ala Ser Ala Gln Thr Leu Asn Cys Ile Tyr Asp Gln Asp Ile Asp  
 20 65 70 75 80  
 Tyr Glu Met Leu Arg Thr Arg Ala Arg Pro Ile Pro Ala Gly Lys Val  
 85 90 95  
 Gln Pro Arg His Ala Leu Ile Phe Ala Leu Ala Leu Gly Val Leu Ser  
 100 105 110  
 25 Phe Ala Leu Leu Ala Thr Phe Val Asn Val Leu Ser Gly Cys Leu Ala  
 115 120 125  
 Leu Ser Gly Ile Val Phe Tyr Met Leu Val Tyr Thr His Trp Leu Lys  
 130 135 140  
 Arg His Thr Ala Gln Asn Ile Val Ile Gly Gly Ala Ala Gly Ser Ile  
 30 145 150 155 160  
 Pro Pro Leu Val Gly Trp Ala Ala Val Thr Gly Asp Leu Ser Trp Thr  
 165 170 175  
 Pro Trp Val Leu Phe Ala Leu Ile Phe Leu Trp Thr Pro Pro His Phe  
 180 185 190  
 35 Trp Ala Leu Ala Leu Met Ile Lys Asp Asp Tyr Ala Gln Val Asn Val  
 195 200 205  
 Pro Met Leu Pro Val Ile Ala Gly Glu Glu Lys Thr Val Ser Gln Ile  
 210 215 220  
 Trp Tyr Tyr Ser Leu Leu Val Val Pro Phe Ser Leu Leu Leu Val Tyr

25

165 170 175  
 Leu Gly Ala Ser Tyr Ile Ala Leu Pro Trp Trp Ala Gly His Ala Leu  
 180 185 190  
 Phe Gly Thr Leu Asn Pro Thr Ile Met Val Leu Thr Leu Ile Tyr Ser  
 5 195 200 205  
 Leu Ala Gly Leu Gly Ile Ala Val Val Asn Asp Phe Lys Ser Val Glu  
 210 215 220  
 Gly Asp Arg Gln Leu Gly Leu Lys Ser Leu Pro Val Met Phe Gly Ile  
 225 230 235 240  
 10 Gly Thr Ala Ala Trp Ile Cys Val Ile Met Ile Asp Val Phe Gln Ala  
 245 250 255  
 Gly Ile Ala Gly Tyr Leu Ile Tyr Val His Gln Gln Leu Tyr Ala Thr  
 260 265 270  
 Ile Val Leu Leu Leu Leu Ile Pro Gln Ile Thr Phe Gln Asp Met Tyr  
 15 275 280 285  
 Phe Leu Arg Asn Pro Leu Glu Asn Asp Val Lys Tyr Gln Ala Ser Ala  
 290 295 300  
 Gln Pro Phe Leu Val Phe Gly Met Leu Ala Thr Gly Leu Ala Leu Gly  
 305 310 315 320  
 20 His Ala Gly Ile

&lt;210&gt; 35

&lt;211&gt; 307

25 &lt;212&gt; PRT

&lt;213&gt; Synechocystis sp

&lt;400&gt; 35

30 Met Thr Glu Ser Ser Pro Leu Ala Pro Ser Thr Ala Pro Ala Thr Arg  
 1 5 10 15  
 Lys Leu Trp Leu Ala Ala Ile Lys Pro Pro Met Tyr Thr Val Ala Val  
 20 25 30  
 Val Pro Ile Thr Val Gly Ser Ala Val Ala Tyr Gly Leu Thr Gly Gln  
 35 40 45  
 35 Trp His Gly Asp Val Phe Thr Ile Phe Leu Leu Ser Ala Ile Ala Ile  
 50 55 60  
 Ile Ala Trp Ile Asn Leu Ser Asn Asp Val Phe Asp Ser Asp Thr Gly  
 65 70 75 80  
 Ile Asp Val Arg Lys Ala His Ser Val Val Asn Leu Thr Gly Asn Arg

	85	90	95
	Asn Leu Val Phe Leu Ile Ser Asn Phe Phe Leu Leu Ala Gly Val Leu		
	100	105	110
	Gly Leu Met Ser Met Ser Trp Arg Ala Gln Asp Trp Thr Val Leu Glu		
5	115	120	125
	Leu Ile Gly Val Ala Ile Phe Leu Gly Tyr Thr Tyr Gln Gly Pro Pro		
	130	135	140
	Phe Arg Leu Gly Tyr Leu Gly Leu Gly Glu Leu Ile Cys Leu Ile Thr		
	145	150	155
10	Phe Gly Pro Leu Ala Ile Ala Ala Ala Tyr Tyr Ser Gln Ser Gln Ser		160
	165	170	175
	Phe Ser Trp Asn Leu Leu Thr Pro Ser Val Phe Val Gly Ile Ser Thr		
	180	185	190
	Ala Ile Ile Leu Phe Cys Ser His Phe His Gln Val Glu Asp Asp Leu		
15	195	200	205
	Ala Ala Gly Lys Lys Ser Pro Ile Val Arg Leu Gly Thr Lys Leu Gly		
	210	215	220
	Ser Gln Val Leu Thr Leu Ser Val Val Ser Leu Tyr Leu Ile Thr Ala		
	225	230	235
20	Ile Gly Val Leu Cys His Gln Ala Pro Trp Gln Thr Leu Leu Ile Ile		240
	245	250	255
	Ala Ser Leu Pro Trp Ala Val Gln Leu Ile Arg His Val Gly Gln Tyr		
	260	265	270
	His Asp Gln Pro Glu Gln Val Ser Asn Cys Lys Phe Ile Ala Val Asn		
25	275	280	285
	Leu His Phe Phe Ser Gly Met Leu Met Ala Ala Gly Tyr Gly Trp Ala		
	290	295	300
	Gly Leu Gly		
	305		
30			
	<210> 36		
	<211> 927		
	<212> DNA		
	<213> Synechocystis sp		
35			
	<400> 36		
	atggcaacta tccaagcttt ttggcgcttc tcccgccccc ataccatcat tggtacaact		60
	ctgagcgtct gggctgtgta tctgttaact attctcgggg atggaaactc agttaactcc		120
	cctgcttccc tggatttagt gttcggcgct tggtggcct gcctgttggg taatgtgtac		180

attgtcggcc tcaaccaatt gtgggatgtg gacattgacc gcatcaataa gccgaatttg 240  
 cccctagcta acggagattt ttctatcgcc cagggccggtt ggattgtggg actttgtggc 300  
 gttgcttcct tggcgatcgc ctggggatta gggctatggc tggggctaac ggtgggcatt 360  
 agtttgatta ttggcacggc ctattcgggtg ccgccagtga ggttaaagcg cttttccctg 420  
 5 ctggcgggccc tgtgtattct gacggtgcgg ggaattgtgg ttaacttggg cttattttta 480  
 ttttttagaa ttggtttagg ttatcccccc actttaataa ccccatctg ggttttgact 540  
 ttatttatct tagttttcac cgtggcgatc gccattttta aagatgtgcc agatatggaa 600  
 ggcgatcggc aatttaagat tcaaacttta actttgcaaa tcggcaaaca aaacgttttt 660  
 cggggaacct taattttact cactggttgt tatttagcca tggcaatctg gggcttatgg 720  
 10 gcggctatgc ctttaaatac tgctttcttg attgtttccc atttgtgctt attagcctta 780  
 ctctggtggc ggagtcgaga tgtacactta gaaagcaaaa ccgaaattgc tagtttttat 840  
 cagtttattt ggaagctatt tttcttagag tacttgctgt atcccttggc tctgtggtta 900  
 cctaattttt ctaatactat ttttttag 927

15 <210> 37  
 <211> 308  
 <212> PRT  
 <213> Synechocystis sp

20 <400> 37  
 Met Ala Thr Ile Gln Ala Phe Trp Arg Phe Ser Arg Pro His Thr Ile  
 1 5 10 15  
 Ile Gly Thr Thr Leu Ser Val Trp Ala Val Tyr Leu Leu Thr Ile Leu  
 20 25 30  
 25 Gly Asp Gly Asn Ser Val Asn Ser Pro Ala Ser Leu Asp Leu Val Phe  
 35 40 45  
 Gly Ala Trp Leu Ala Cys Leu Leu Gly Asn Val Tyr Ile Val Gly Leu  
 50 55 60  
 Asn Gln Leu Trp Asp Val Asp Ile Asp Arg Ile Asn Lys Pro Asn Leu  
 30 65 70 75 80  
 Pro Leu Ala Asn Gly Asp Phe Ser Ile Ala Gln Gly Arg Trp Ile Val  
 85 90 95  
 Gly Leu Cys Gly Val Ala Ser Leu Ala Ile Ala Trp Gly Leu Gly Leu  
 100 105 110  
 35 Trp Leu Gly Leu Thr Val Gly Ile Ser Leu Ile Ile Gly Thr Ala Tyr  
 115 120 125  
 Ser Val Pro Pro Val Arg Leu Lys Arg Phe Ser Leu Leu Ala Ala Leu  
 130 135 140  
 Cys Ile Leu Thr Val Arg Gly Ile Val Val Asn Leu Gly Leu Phe Leu

145                      150                      155                      160  
 Phe Phe Arg Ile Gly Leu Gly Tyr Pro Pro Thr Leu Ile Thr Pro Ile  
                          165                      170                      175  
 Trp Val Leu Thr Leu Phe Ile Leu Val Phe Thr Val Ala Ile Ala Ile  
 5                      180                      185                      190  
 Phe Lys Asp Val Pro Asp Met Glu Gly Asp Arg Gln Phe Lys Ile Gln  
                          195                      200                      205  
 Thr Leu Thr Leu Gln Ile Gly Lys Gln Asn Val Phe Arg Gly Thr Leu  
                          210                      215                      220  
 10 Ile Leu Leu Thr Gly Cys Tyr Leu Ala Met Ala Ile Trp Gly Leu Trp  
                          225                      230                      235                      240  
 Ala Ala Met Pro Leu Asn Thr Ala Phe Leu Ile Val Ser His Leu Cys  
                          245                      250                      255  
 Leu Leu Ala Leu Leu Trp Trp Arg Ser Arg Asp Val His Leu Glu Ser  
 15                      260                      265                      270  
 Lys Thr Glu Ile Ala Ser Phe Tyr Gln Phe Ile Trp Lys Leu Phe Phe  
                          275                      280                      285  
 Leu Glu Tyr Leu Leu Tyr Pro Leu Ala Leu Trp Leu Pro Asn Phe Ser  
                          290                      295                      300  
 20 Asn Thr Ile Phe  
                          305

&lt;210&gt; 38

&lt;211&gt; 1092

25 &lt;212&gt; DNA

&lt;213&gt; Synechocystis sp

&lt;400&gt; 38

atgaaatttc cgccccacag tggttaccat tggcaaggtc aatcaccttt ctttgaagggt 60  
 30 tgggtacgtgc gcctgctttt gccccaatcc ggggaaagtt ttgcttttat gtactccatc 120  
 gaaaatcctg ctagcgatca tcattacggc ggcgggtgctg tgcaaatttt agggccggct 180  
 acgaaaaaac aagaaaatca ggaagaccaa cttgtttggc ggacatttcc ctcggtaaaa 240  
 aaattttggg ccagtcctcg ccagtttgcc ctagggcatt ggggaaaatg tagggataac 300  
 aggcaggcga aaccctact ctccgaagaa ttttttgcca cgggtcaagga aggttatcaa 360  
 35 atccatcaaa atcagcacca aggacaaatc attcatggcg atcgccattg tcgttggcag 420  
 ttcaccgtag aaccggaagt aacttggggg agtcctaacc gatttcctcg ggctacagcg 480  
 gggttggttt cttttttacc cttgtttgat cccggttggc aaattctttt agcccaaggt 540  
 agagcgcacg gctggctgaa atggcagagg gaacagtatg aatttgacca cgccctagtt 600  
 tatgccgaaa aaaattgggg tcactccttt ccctcccgtt ggttttggct ccaagcaaat 660

	tatttttcctg accatccagg actgagcgtc actgccgctg gcggggaacg gattgttctt	720
	ggtcgccccg aagaggtagc tttaattggc ttacatcacc aaggtaattt ttacgaattt	780
	ggcccggggc atggcacagt cacttggcaa gtagctccct ggggccgttg gcaattaaaa	840
	gccagcaatg ataggtattg ggtcaagttg tccggaaaaa cagataaaaa aggcagttta	900
5	gtccacactc ccaccgcca gggcttacia ctcaactgcc gagataccac taggggctat	960
	ttgtatttgc aattgggatc tgtgggtcac ggcctgatag tgcaagggga aacggacacc	1020
	gcggggctag aagttggagg tgattggggt ttaacagagg aaaatttgag caaaaaaca	1080
	gtgccattct ga	1092
10	<210> 39	
	<211> 363	
	<212> PRT	
	<213> Synechocystis sp	
15	<400> 39	
	Met Lys Phe Pro Pro His Ser Gly Tyr His Trp Gln Gly Gln Ser Pro	
	1 5 10 15	
	Phe Phe Glu Gly Trp Tyr Val Arg Leu Leu Pro Gln Ser Gly Glu	
	20 25 30	
20	Ser Phe Ala Phe Met Tyr Ser Ile Glu Asn Pro Ala Ser Asp His His	
	35 40 45	
	Tyr Gly Gly Gly Ala Val Gln Ile Leu Gly Pro Ala Thr Lys Lys Gln	
	50 55 60	
	Glu Asn Gln Glu Asp Gln Leu Val Trp Arg Thr Phe Pro Ser Val Lys	
25	65 70 75 80	
	Lys Phe Trp Ala Ser Pro Arg Gln Phe Ala Leu Gly His Trp Gly Lys	
	85 90 95	
	Cys Arg Asp Asn Arg Gln Ala Lys Pro Leu Leu Ser Glu Glu Phe Phe	
	100 105 110	
30	Ala Thr Val Lys Glu Gly Tyr Gln Ile His Gln Asn Gln His Gln Gly	
	115 120 125	
	Gln Ile Ile His Gly Asp Arg His Cys Arg Trp Gln Phe Thr Val Glu	
	130 135 140	
	Pro Glu Val Thr Trp Gly Ser Pro Asn Arg Phe Pro Arg Ala Thr Ala	
35	145 150 155 160	
	Gly Trp Leu Ser Phe Leu Pro Leu Phe Asp Pro Gly Trp Gln Ile Leu	
	165 170 175	
	Leu Ala Gln Gly Arg Ala His Gly Trp Leu Lys Trp Gln Arg Glu Gln	
	180 185 190	



Tyr Glu Phe Asp His Ala Leu Val Tyr Ala Glu Lys Asn Trp Gly His  
 .195 200 205  
 Ser Phe Pro Ser Arg Trp Phe Trp Leu Gln Ala Asn Tyr Phe Pro Asp  
 210 215 220  
 5 His Pro Gly Leu Ser Val Thr Ala Ala Gly Gly Glu Arg Ile Val Leu  
 225 230 235 240  
 Gly Arg Pro Glu Glu Val Ala Leu Ile Gly Leu His His Gln Gly Asn  
 245 250 255  
 Phe Tyr Glu Phe Gly Pro Gly His Gly Thr Val Thr Trp Gln Val Ala  
 10 260 265 270  
 Pro Trp Gly Arg Trp Gln Leu Lys Ala Ser Asn Asp Arg Tyr Trp Val  
 275 280 285  
 Lys Leu Ser Gly Lys Thr Asp Lys Lys Gly Ser Leu Val His Thr Pro  
 290 295 300  
 15 Thr Ala Gln Gly Leu Gln Leu Asn Cys Arg Asp Thr Thr Arg Gly Tyr  
 305 310 315 320  
 Leu Tyr Leu Gln Leu Gly Ser Val Gly His Gly Leu Ile Val Gln Gly  
 325 330 335  
 Glu Thr Asp Thr Ala Gly Leu Glu Val Gly Gly Asp Trp Gly Leu Thr  
 20 340 345 350  
 Glu Glu Asn Leu Ser Lys Lys Thr Val Pro Phe  
 355 360

<210> 40  
 25 <211> 56  
 <212> DNA  
 <213> Artificial Sequence

<400> 40  
 30 cgcgatttaa atggcgcgcc ctgcaggcgg ccgcctgcag ggcgcgccat ttaaat 56

<210> 41  
 <211> 32  
 <212> DNA  
 35 <213> Artificial Sequence

<400> 41  
 tcgaggatcc gcggccgcaa gcttcctgca gg

32

<210> 42  
<211> 32  
<212> DNA  
<213> Artificial Sequence  
5  
<400> 42  
tcgacctgca ggaagcttgc ggccgcggat cc 32

<210> 43  
10 <211> 32  
<212> DNA  
<213> Artificial Sequence  
  
<400> 43  
15 tcgacctgca ggaagcttgc ggccgcggat cc 32

<210> 44  
<211> 32  
<212> DNA  
20 <213> Artificial Sequence  
  
<400> 44  
tcgaggatcc gcggccgcaa gcttcctgca gg 32

25 <210> 45  
<211> 36  
<212> DNA  
<213> Artificial Sequence  
  
<400> 45  
30 tcgaggatcc gcggccgcaa gcttcctgca ggagct 36

<210> 46  
<211> 28  
35 <212> DNA  
<213> Artificial Sequence  
  
<400> 46  
cctgcaggaa gcttgccggcc gcggatcc 28

<210> 47  
<211> 36  
<212> DNA  
5 <213> Artificial Sequence

<400> 47  
tcgacctgca ggaagcttgc ggccgcggat ccagct 36

10 <210> 48  
<211> 28  
<212> DNA  
<213> Artificial Sequence

15 <400> 48  
ggatccgcgg ccgcaagctt cctgcagg 28

<210> 49  
<211> 39  
20 <212> DNA  
<213> Artificial Sequence

<400> 49  
gatcacctgc aggaagcttg cggccgcgga tccaatgca 39

25 <210> 50  
<211> 31  
<212> DNA  
<213> Artificial Sequence

30 <400> 50  
ttggatccgc ggccgcaagc ttcctgcagg t 31

<210> 51  
35 <211> 41  
<212> DNA  
<213> Artificial Sequence

<400> 51

ggatccgcgg ccgcacaatg gagtctctgc tctctagttc t 41

<210> 52  
<211> 38  
5 <212> DNA  
<213> Artificial Sequence

<400> 52  
10 ggatcctgca ggtcacttca aaaaaggtaa cagcaagt 38

<210> 53  
<211> 45  
<212> DNA  
<213> Artificial Sequence  
15

<400> 53  
ggatccgcgg ccgcacaatg gcgttttttg ggctctcccg tgttt 45

<210> 54  
20 <211> 40  
<212> DNA  
<213> Artificial Sequence

<400> 54  
25 ggatcctgca ggttattgaa aacttcttcc aagtacaact 40

<210> 55  
<211> 38  
<212> DNA  
30 <213> Artificial Sequence

<400> 55  
ggatccgcgg ccgcacaatg tggcgaagat ctgttggt 38

35 <210> 56  
<211> 37  
<212> DNA  
<213> Artificial Sequence

<400> 56  
ggatcctgca ggtcatggag agtagaagga aggagct 37

<210> 57  
5 <211> 50  
<212> DNA  
<213> Artificial Sequence

<400> 57  
10 ggatccgcgg ccgcacaatg gtacttgccg aggttccaaa gcttgccctct 50

<210> 58  
<211> 38  
<212> DNA  
15 <213> Artificial Sequence

<400> 58  
ggatcctgca ggtcacttgt ttctggtgat gactctat 38

<210> 59  
20 <211> 38  
<212> DNA  
<213> Artificial Sequence

<400> 59  
25 ggatccgcgg ccgcacaatg acttcgattc tcaacact 38

<210> 60  
<211> 36  
30 <212> DNA  
<213> Artificial Sequence

<400> 60  
35 ggatcctgca ggtcagtgtt gcgatgctaa tgccgt 36

<210> 61  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<400> 61  
taatgtgtac attgtcggcc tc 22

5 <210> 62  
<211> 60  
<212> DNA  
<213> Artificial Sequence

10 <400> 62  
gcaatgtaac atcagagatt ttgagacaca acgtggcttt ccacaattcc ccgcaccgtc 60

<210> 63  
<211> 22

15 <212> DNA  
<213> Artificial Sequence

<400> 63  
aggctaataa gcacaaatgg ga 22

20 <210> 64  
<211> 63  
<212> DNA  
<213> Artificial Sequence

25 <400> 64  
ggtatgagtc agcaacacct tcttcacgag gcagacctca gcggaattgg tttaggttat 60  
ccc 63

30 <210> 65  
<211> 26  
<212> DNA  
<213> Artificial Sequence

35 <400> 65  
ggatccatgg ttgcccaaac cccatc 26

<210> 66  
<211> 61

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;400&gt; 66

5 gcaatgtaac atcagagatt ttgagacaca acgtggcttt gggtaagcaa caatgaccgg 60  
c 61

&lt;210&gt; 67

&lt;211&gt; 25

10 &lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;400&gt; 67

15 gaattctcaa agccagccca gtaac 25

&lt;210&gt; 68

&lt;211&gt; 63

&lt;212&gt; DNA

20 <213> Artificial Sequence

&lt;400&gt; 68

ggatatgagtc agcaacacct ttttcacgag gcagacctca gcgggtgcga aaagggtttt 60  
ccc 63

25 &lt;210&gt; 69

&lt;211&gt; 23

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

30 &lt;400&gt; 69

ccagtggttt aggctgtgtg gtc 23

&lt;210&gt; 70

&lt;211&gt; 21

35 &lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;400&gt; 70

ctgagttgga tgtattggat c 21

<210> 71  
<211> 28  
<212> DNA  
5 <213> Artificial Sequence

<400> 71  
ggatccatgg ttacttcgac aaaaatcc 28

10 <210> 72  
<211> 60  
<212> DNA  
<213> Artificial Sequence

15 <400> 72  
gcaatgtaac atcagagatt ttgagacaca acgtggcttt gctaggcaac cgcttagtac 60

<210> 73  
<211> 28  
20 <212> DNA  
<213> Artificial Sequence

<400> 73  
gaattcttaa cccaacagta aagttccc 28

25 <210> 74  
<211> 63  
<212> DNA  
<213> Artificial Sequence

30 <400> 74  
ggtatgagtc agcaacacct tcttcacgag gcagacctca ggcgcggcat tgtcttttac 60  
atg 63

35 <210> 75  
<211> 20  
<212> DNA  
<213> Artificial Sequence



<400> 75  
ggaacccttg cagccgcttc 20

<210> 76  
5 <211> 22  
<212> DNA  
<213> Artificial Sequence

<400> 76  
10 gtatgcccaa ctggtgcaga gg 22

<210> 77  
<211> 28  
<212> DNA  
15 <213> Artificial Sequence

<400> 77  
ggatccatgt ctgacacaca aaataccg 28

20 <210> 78  
<211> 62  
<212> DNA  
<213> Artificial Sequence

25 <400> 78  
gcaatgtaac atcagagatt ttgagacaca acgtggcttt cgccaatacc agccaccaac 60  
ag 62

<210> 79  
30 <211> 27  
<212> DNA  
<213> Artificial Sequence

<400> 79  
35 gaattctcaa atccccgcat ggcctag 27

<210> 80  
<211> 65  
<212> DNA

<213> Artificial Sequence

<400> 80  
ggatatgagtc agcaacacct tcttcacgag gcagacctca gcggcctacg gcttggacgt 60  
5 gtggg 65

<210> 81  
<211> 21  
<212> DNA  
10 <213> Artificial Sequence

<400> 81  
cacttggatt cccctgatct g 21

15 <210> 82  
<211> 21  
<212> DNA  
<213> Artificial Sequence

20 <400> 82  
gcaatacccg cttggaaaac g 21

<210> 83  
<211> 29  
25 <212> DNA  
<213> Artificial Sequence

<400> 83  
ggatccatga ccgaatcttc gccctagc 29  
30

<210> 84  
<211> 61  
<212> DNA  
<213> Artificial Sequence  
35

<400> 84  
gcaatgtaac atcagagatt ttgagacaca acgtggcttt caatcctagg tagccgaggc 60  
g 61

<210> 85  
<211> 27  
<212> DNA  
<213> Artificial Sequence  
5  
<400> 85  
gaattccttag cccaggccag cccagcc  
27  
  
<210> 86  
10 <211> 66  
<212> DNA  
<213> Artificial Sequence  
  
<400> 86  
15 ggtatgagtc agcaacacct tcttcacgag gcagacctca gcggggaatt gatttgttta  
attacc  
60  
66  
  
<210> 87  
<211> 21  
20 <212> DNA  
<213> Artificial Sequence  
  
<400> 87  
gcgatcgcca ttatcgcttg g  
25  
21  
  
<210> 88  
<211> 24  
<212> DNA  
<213> Artificial Sequence  
30  
  
<400> 88  
gcagactggc aattatcagt aacg  
24  
  
<210> 89  
35 <211> 25  
<212> DNA  
<213> Artificial Sequence  
  
<400> 89

ccatggattc gagtaaagtt gtcgc

25

&lt;210&gt; 90

&lt;211&gt; 0

5 &lt;213&gt; Artificial Sequence

&lt;400&gt; 90

gaattcactt caaaaaaggt aacag

10 &lt;210&gt; 91

&lt;211&gt; 4550

&lt;212&gt; DNA

&lt;213&gt; Arabidopsis sp

15 &lt;400&gt; 91

	attttacacc aatttgatca cttaactaaa ttaattaaat tagatgatta tcccaccata	60
	tttttgagca ttaaaccata aaaccatagt tataagtaac tgttttaate gaatatgact	120
	cgattaagat taggaaaaat ttataaccgg taattaagaa aacattaacc gtagtaaccg	180
	taaatgcoga ttccctccctt gtctaaaaga cagaaaacat atattttatt ttgcccata	240
20	tgtttcactc tatttaattt caggcacaat acttttggtt ggtaacaaaa ctaaaaagga	300
	caacacgtga tacttttcct cgtccgtcag tcagattttt tttaaactag aaacaagtgg	360
	caaactaca ccacattttt tgcttaatct attaacttgt aagtttttaa ttccataaaa	420
	agtctaacta attcttctaa tataagtaca ttccctaaat ttcccaaaaa gtcaaattaa	480
	taattttcaa aatctaactt aaatatctaa taattcaaaa tcattaataaa gacacgcaac	540
25	aatgacacca attaatcacc ctgcaccac acaattctac agttctcatg cttaaccata	600
	ttttttgtct tctgttcctt caaaatcatt tctttctctt ctttgattcc caaagatcac	660
	ttctttgtct ttgatttttg attttttttc tctctggcgt gaaggaagaa gctttatttc	720
	atggagtctc tgctctctag ttcttctctt gtttccgctg gtaaactctg tccttttctg	780
	gtttcagggt ttatttggtg tttaggtttc gtttttggtg ttcagaacca tacaaaaagt	840
30	ttgaactttt ctgaatataa aataaggaaa aagtttcgat ttttataatg aattgtttac	900
	tagatcgaag taggtgacaa aggttattgt gtggagaagc ataatttctg ggcttgactt	960
	tgaattttgt ttctcatgca tgcaacttat caatcagctg gtgggttttg ttggaagaag	1020
	cagaatctaa agctccactc tttatcagggt tcgttaggggt tttatgggtt tttgaaatta	1080
	aatactcaat catcttagtc tcattattct attgggtgaa tcacattttc taatttgga	1140
35	tttatgagac aatgtatggt ggacttaggt gaagttcttc tctttggtta tagttgaagt	1200
	gttactgatg ttgtttagct ctttacacca atatatacac ccaattttgc agaaatccga	1260
	gttctgcgtt gtgattcgag taaagttgtc gcaaaaccga agtttaggaa caatcttggt	1320
	aggcctgatg gtcaaggatc ttcattgttg ttgtatccaa aacataagtc gagatttcgg	1380
	gttaatgcca ctgcgggtca gcctgaggct ttcgactcga atagcaaaca gaagtctttt	1440

	agagactcgt tagatgcgtt ttacagggtt tctaggcctc atacagttat tggcacagtt	1500
	aagtttctct ttaaaaatgt aactctttta aaacgcaatc tttcagggtt ttcaaggaga	1560
	taacattagc tctgtgattg gatttgcagg tgcttagcat tttatctgta tctttcttag	1620
	cagtagagaa ggtttctgat atatctcctt tacttttcac tggcatcttg gaggtaatga	1680
5	atatataaca cataatgacc gatgaagaag atacattttt ttcgtctctc tgtttaaaca	1740
	attgggtttt gttttcaggc tgttggtgca gctctcatga tgaacattta catagttggg	1800
	ctaaatcagt tgtctgatgt tgaaatagat aaggtaacat gcaaattttc ttcatatgag	1860
	ttcgagagac tgatgagatt aatagcagct agtgccctaga tcatctctat gtgggttttt	1920
	gcaggttaac aagccctatc ttccattggc atcaggagaa tattctgtta acaccggcat	1980
10	tgcaatagta gcttccttct ccatcatggg atgggtgccat tttcacaaaa tttcaacttt	2040
	tagaattcta taagttactg aaatagtttg ttataaatcg ttatagagtt tctggcttgg	2100
	gtggattgtt ggttcattggc cattgttctg ggctcttttt gtgagtttca tgctcggtac	2160
	tgcatactct atcaatgtaa gtaagtttct caataactaga atttggctca aatcaaaatc	2220
	tgcagtttct agtttttaggt taatgaggtt ttaataactt acttctacta caaacagttg	2280
15	ccacttttac ggtggaaaag atttgcattg gttgcagcaa tgtgtatcct cgctgtccga	2340
	gctattattg ttcaaatcgc cttttatcta catattcagg tactaaacca ttttccttat	2400
	gttttgtagt tgttttcac c aaatcactt ttatattact aaagctgtga aactttgttg	2460
	cagacacatg tgtttggaag accaatcttg ttacttaggc ctcttatttt cgccactgcg	2520
	tttatgagct ttttctctgt cgttattgca ttgtttaagg taaacaaaga tggaaaaaga	2580
20	ttaaatctat gtatacttaa agtaaagcat tctactgtta ttgatgagaa gttttctttt	2640
	ttgggttgat gcaggatata cctgatatcg aaggggataa gatattcgga atccgatcat	2700
	tctctgtaac tctgggtcag aaacgggtac gatatctaaa ctaaagaaat tgttttgact	2760
	caagtgttgg attaagatta cagaagaaag aaaactgttt ttgtttcttg caaaattcag	2820
	gtgttttgga catgtgttac actacttcaa atggccttac ctgttgcaat tctagttgga	2880
25	gccacatctc cattcatatg gagcaaagtc atctcggtaa caatctttct ttacccatcg	2940
	aaaactcgc aattcatcgt ttgagtggta ctggtttcat tttgttccgt tctgttgatt	3000
	ttttttcagg ttgtgggtca tgttatactc gcaacaactt tgtgggctcg agctaagtcc	3060
	gttgatctga gtagcaaaac cgaaataact tcatgttata tgttcatatg gaaggtaga	3120
	ttcgtttata aatagagtct ttactgcctt tttatgcgct ccaatttgga attaaaatag	3180
30	cctttcagtt tcatcgaatc accattatac tgataaatte tcatctctgc atcagctctt	3240
	ttatgcagag tacttgctgt tacctttttt gaagtgactg acattagaag agaagaagat	3300
	ggagataaaa gaataagtca tcaactatgct tctgttttta ttacaagttc atgaaattag	3360
	gtagtgaact agtgaattag agttttatct tgaaacatgg cagactgcaa aaatatgtca	3420
	aagatatgaa tttctgttgg gtaaagaagt ctctgcttgg gcaaaatctt aaggttcggg	3480
35	gtgttgatat aatgctaage gaagaaatcg attctatgta gaaatttccg aaactatgtg	3540
	taaacatgtc agaacatctc cattctatat cttcttctgc aagaaagctc tgtttttatc	3600
	acctaaactc tttatctctg ttagttaaag atatgtatat gtacgtgact acattttttt	3660
	gttgatgtaa tttgcagaac gtatggattt ttgttagaaa gcatgagttc gaaagtatat	3720
	gtttatatat atggataatt cagacctaac gtcgaagctc acaagcataa attcactact	3780

	atagtttgct ctgtaataga tagttccatt gatgtcttga aactgtacgt aactgcctgg	3840
	gcgttttgtg gttgatactg actactgagt gttctttgtg agtggtgtaa gtatacaaga	3900
	agaagaatat aggctcacgg gaacgactgt ggtggaagat gaaatggaga tcatcacgta	3960
	gcggctttgc caaagaccga gtcacgatcg agtctatgaa gtctttacag ctgctgatta	4020
5	tgattgacca ttgcttagag acgcattgga atcttactag ggacttgcct gggagtctct	4080
	tcaagtacgt gtcagatcat acgatgtagg agatttcacg gctttgatgt gtttgtttgg	4140
	agtcacaatg cttaatgggc ttattggccc aataatagct agctcttttg ctttagccgt	4200
	ttcgtttgtc ccctgggtgg gagtattatt agggatgggt gtgaccaaag tcaccagacc	4260
	tagagtgaat ctagtagagt cctagaccat ggtccatggc ttttatttgt aatttgaaaa	4320
10	atgaacaatt ctttttgtaa ggaaaacttt tatatagtag acgtttacta tatagaaact	4380
	agttgaacta acttcgtgca attgcataat aatgggtgtga aatagaggggt gcaaaaactca	4440
	ataaacattt cgacgtacca agagttcgaa acaataagca aaatagattt ttttgcttca	4500
	gactaatttg tacaatgaat ggttaataaa ccattgaagc ttttattaat	4550
15	<210> 92	
	<211> 4450	
	<212> DNA	
	<213> Arabidopsis sp	
20	<400> 92	
	tttaggttac aaaatcaatg atattgcgta tgtcaactat aaaagccaaa agtaaagcct	60
	cttgtttgac cagaagggtca tgatcattgt atacatacag ccaaactacc tcctggaaga	120
	aaagacatgg atcccaaaca acaacaatag cttctttttac aagaaccagt agtaactagt	180
	cactaatcta aaagagttaa gtttcagctt ttctggcaat ggctccttga tcatttcaat	240
25	cctgaaggag acccactttg tagcaagacc atgtcctctg tttcacttac agtgtgtctc	300
	aaaagtctac ttcaattctt catatatagg ttcttcacac tacagcttca tcctcattcg	360
	ttgacagaga gagagtcttt attgaaaact tcttccaagt acaactccac taaatataat	420
	agcaccaaac cacttgttcg acacaaatct gtacagatat aaaaacacta ttaggttttc	480
	caaggcaaat cacataattg gattgtgaaa gagtacaaaa gataaaccac aattttcata	540
30	ctttctactg cagtcagcac cagatgataa gtcagctgtc cctatttgcc atcctaactg	600
	tcctgatgca gcggccagtg atgcgtaata ttgccaccct taatcattag agcgagaaac	660
	aaaaagaatc aaaagacagt aaatggaatt aggaatcaca aatgagtcct tgtaaagttt	720
	attgagtacc gagatctgca ctgaatccag aaagtgcaag aaaacctatg gatgctgtgc	780
	caaatccagt taaccaaaagc tttgtattat caccgaatct aagggtgtgt gacttaacac	840
35	caacttttac atcatcttct ttgtcctgga gacacaatat attagacatt agtccatgga	900
	aaaaaaatga tttaacctag aatatctcaa aattacttgc ataaaaactg aacttgagct	960
	gaaattttgg gttcgtagct tgtggcatat actatttcat tttcaatggg ccacaaagg	1020
	aactttcttt tctcacttct gttgcaaagc ggaagacttt tatggggcta actcttcact	1080
	taaagtatag aaatcagatg gaaaagggtg gagatcaggg taattttctt ctttatgatt	1140

	gacaaaagtc gaacatcgaa atggatgcat ttgcatgaga catgaaacaa aagctgaaaa	1200
	agaaatctgt ggtggtgaag ctagaaaaag aaaacaaagc aagcaatatg cacacattga	1260
	gattaactac tttgctactg gtcataatca aatagatttt gaagctaaaa aataaaaagt	1320
	gaatatacct gatgtgcata aatagtatca taaacaaggg tccagcagac tccggagaga	1380
5	tagagaggga gtacaataga tgggtgctatg cttcctttta ctgcagtcca tcctaacaat	1440
	gctccccagt ttatgggtcaa acctaaaaag gcttgaggct gcaattataa aaacgaatca	1500
	atcataagaa aatcagaaaa tatataatgt ctaactttga gaagccagaa tagattttaa	1560
	ttacccaaaa tgtaaacctc ttcataagtg ggtaggaaaa gacaagtaac aaagatgaag	1620
	cccctaaaac acggctgcag aatatacata ctgaaatgag ctcaagtaga aaagaatttg	1680
10	atcacaaaac taaagacaag acctgagaac atatcttcag aatttgggccc aactacataa	1740
	gggtgaacca tatgtgtatg tgaattttta aacaaacact tgcaaatagc cgacttttagg	1800
	gcaagtaaaa aatccaaaca aacctgtaat tgttaagttg gagaagaatc cctaagccta	1860
	aaagcaactg cagcccgaga aatccaatcc cttgaaatgg tgtcaaaaga ccaactggcga	1920
	taggtcttag ttttgtacga tcaacctgga tataaaagaa atttgtaaga caacataatc	1980
15	taaaacaaaa caaccataca aaatcttgag ctttacatac aagcaacca tctttgttta	2040
	tggaagaatg aatccagtta catgaatgct gtgtatctac cctaactact aaacacatat	2100
	ttcaatcgaa aaacatattc caccttcacc atatctaaca cctgaagtct ttcacttttt	2160
	gaacgaagtc atcagaacat gcagataagc tattacccaa aacagagata tgactggaaa	2220
	tgttgtcgta aattgatcca acatagaaaa atcaagacca gttccagatg tcaaagcaat	2280
20	aacactttcc caccatgggtt acagaaacca tagttacaca aaacatgttt cctaaaccaa	2340
	catactaaag ggatatataa atttgacatc actttatcac cataccataa gatagcttaa	2400
	aaacaaactg acctttgtat ctatgtcctg atcaagcaga tcatttatag tacaaccagc	2460
	acctctaaga agtaatgctc cgcaaccaa taaagccata tattttaaacc ttggaaggct	2520
	tccaggatca gcagccaacg caatcgacct atacaacaat gatggagatt cagagtatcg	2580
25	atctatttac atagctctgg aactagatcc atgacgaaac atggaacatc gttataatat	2640
	ctaaagactt ccaaacagat tcctgagtaa gaaaccagc ggaactatag tactgtaaca	2700
	tatataaaat caaagaaaac tcaggtttat agcattatcc aatcctgatt tctgccaatc	2760
	cttaaccact ctcccatgct atcaaaaacc tcagctcaag atcatactac ctaattgcct	2820
	atgagctctt gggaagatca ttatggattt gataactgaa aaaagtaaca gagaaatagc	2880
30	agactgcaag aactactcca aacttctcca ctgatatgta tgtagtctaa caataataaa	2940
	cagacataaa ttctttttatc aagcttcaag agcaagtttag tcagaaaaca tcacagccaa	3000
	accaaccagg aaaacacata actttatcac ataaaactaa atttaagtga atctgactta	3060
	acataaacca tcttttggga cgaaaggaaa ctatataaac atgcagtctt tctttccctc	3120
	agctattctt tcggatggat tataatgaat ctcaaaagtg aaatgtcttg attctcagct	3180
35	acattactca aaggcgaaga taaacttacc acatacaagg ccacgcaagc aaccaagttc	3240
	caatgggttt atccaatcga gcaagcttag cataacctct aacttctctt ggtaaataca	3300
	aatctatcca agaagcttcc ttaacaacaa caccatcact cttctcctta tcatctttct	3360
	tcggctttcc ctccaaaacc gaagaagacg acgacattcc acaaattaat ctgtaattcc	3420
	aaccaacacc aaaaaacttc tctgatgca attctcttcc tttactccat acttggtaat	3480

tatcattcca tgaaggataa cacttagtga aaggatttgt gtaatgggta gtcacaggat 3540  
 tggacaagga tttatgttgt gattgcaaaa gagcagagga agaagatgga gttacggaga 3600  
 cggaagattt caacaaccgt cttgaaacac gggagagccc aaaaaacgcc atctttgaga 3660  
 gaaattgttg cctggaagaa acaaagactt gagatttcaa acgtaagtga attcttacga 3720  
 5 acgaaagcta acttctcaag agaatcagat tagtgattcc tcaaaaacaa aaaaaactat 3780  
 ctaatttcag tttcgagtga tgaagcctta agaatctaga acctccatgg cgtttctaata 3840  
 ctctcagaga taatcgaatt ccttaaacia tcaaagctta gaaagagaag aacaacaaca 3900  
 acaacaaaaa aaatcagatt aacaaccgac cagagagcaa cgacgacgcc ggcgagaaag 3960  
 agcacgtcgt ctcggagcaa gacttcttct ccagtaacct ggatggatcg ttaatgggcc 4020  
 10 tgtagattat tatatttggg ccgaaacaat tgggtcagca aaaacttggg ggataatgaa 4080  
 gaaacacgta cagtatgcat ttaggctcca aattaattgg ccataataat cgaatcagat 4140  
 aaactaatca acctctacct tacttatttc tcaactgtttt tatttctacc ttagtagttg 4200  
 aagaaacact tttatttatc ttttcgggac ccaaatttga taggatcggg ccattactca 4260  
 tgagcgtcag acacatatta gccttatcag attagtgggg taaggttttt ttaattcggg 4320  
 15 aagaagcaac aatcaatgtc ggagaaatta aagaatctgc atgggcgtgg cgtgatgata 4380  
 tgtgcatatg gagtcagttg ccgatcatat ataactatct ataaactaca tataaagact 4440  
 actaatagat 4450

&lt;210&gt; 93

20 &lt;211&gt; 2850

&lt;212&gt; DNA

&lt;213&gt; Arabidopsis sp

&lt;400&gt; 93

25 aattaaaatt tgagcgggtct aaaccattag accgtttaga gatccctcca acccaaaata 60  
 gtogattttc acgtcttgaa catatatgtg gccttaactct gtgtgggttag taaagacttt 120  
 tatttggtcaa agaaaaacaa ccatggccca acatgttgat actttttattt aattatacaa 180  
 gtacccctga attctctgaa atatatttga ttgaccaga tattaatttt aattatcatt 240  
 tcctgtaaaa gtgaaggagt caccgtgact cgtcgtaatc tgaaaccaat ctgttcatat 300  
 30 gatgaagaag tttctctcgt tctcctccaa cgcgtagaaa attctgacgg cttaacgatg 360  
 tggcgaagat ctgttggtta tcgtttctct tcaagaatct ctgtttcttc ttcgttacca 420  
 aaccctagac tgattccttg gtcccgcgaa ttatgtgccg ttaatagctt ctcccagcct 480  
 ccggtctcga cggaatcaac tgctaagtta gggatcactg gtgttagatc tgatgccaat 540  
 cgagtttttg ccactgctac tgccgcccgt acagctacag ctaccaccgg tgagatttgc 600  
 35 tctagagttg cggctttggc tggattaggc catcactacg ctcgttgtta ttgggagctt 660  
 tctaaagcta aacttaggta tgtgtttact tttcttttct catgaaaaat ctgaaaattt 720  
 ccaattgttg gattcttaaa ttctcatttg ttttatgggt gtagtatgct tgtgggttgc 780  
 acttctggaa ctgggtatat tctgggtacg ggaaatgctg caattagctt cccggggctt 840  
 tgttacacat gtgcaggaac catgatgatt gctgcatctg ctaattcctt gaatcaggtc 900



	attgaaatgt tgagaagttc ataaatttcg aatccttggt gtgtttatgt agttgatctt	960
	gcttgcttat gtttatgtag ttgaaaagtt taaaaatttc taatccttgg tagttgatct	1020
	cgcttgtttg ttttttcatt ttctagattt ttgagataag caatgattct aagatgaaaa	1080
	gaacgatgct aaggccattg ccttcaggac gtattagtgt tccacacgct gttgcatggg	1140
5	ctactattgc tgggtgcttct ggtgcttggt tgttggccag caaggtgaat gtttggtttt	1200
	ttatatgtga tttctttggt ttatgaatgg gtgattgaga gattatggat ctaaactttt	1260
	gcttccacga caaggttatt gcagactaat atgttggctg ctggacttgc atctgccaat	1320
	cttgactttt atgcgtttgt ttatactccg ttgaagcaac ttcacctat caatacatgg	1380
	gttggcgctg ttgttgggtgc tatcccaccc ttgcttgggt aaatttttgt tccttttctt	1440
10	ctttatttta gcagattctg ttttggttga tactgctttt aattcaaaat gtagtcatgg	1500
	ttcaccaatt ctatgcttat ctattttgtg tgttgtcagg tgggcggcag cgtctggtca	1560
	gatttcatac aattcgatga ttcttccagc tgctctttac ttttggcaga tacctcattt	1620
	tatggccctt gcacatctct gccgcaatga ttatgcagct ggagggtgaag accatatggt	1680
	gtcatatgag attagaatgt ctccctccat gtagtggtga tcttgaacta gttcaatttc	1740
15	gtggaatgat cagagtgtcc tagatagtgt cacagcagtc gacatttttag tggctagata	1800
	atgagttctt tccggttagag ataaacattc gcgaacattg tttccagctt ccgcgacca	1860
	acttctgatt ttgtttcttg gtaccttggt ttcagttaca agatgttgtc actctttgat	1920
	ccgtcagggg agagaatagc agcagtggct ctaagggaact gcttttacat gatccctctc	1980
	ggtttcatcg cctatgactg tgagtcttgt agattcatct tttttttgta gtttattgac	2040
20	tgcattgctg tatctgattt ttgctgttcc ttccaatttt tgtgacaggg ggggttaacct	2100
	caagttgggt ttgcctcgaa tcaacacttc tcacactagc aatcgctgca acagcatttt	2160
	cattctaccg agaccggacc atgcataaag caaggaaaat gttccatgcc agtcttctct	2220
	tccttctgt tttcatgtct ggtcttcttc tacaccgtgt ctctaattgat aatcagcaac	2280
	aactcgtaga agaagccgga ttaacaaatt ctgtatctgg tgaagtcaaa actcagaggc	2340
25	gaaagaaacg tgtggctcaa cctccggtgg cttatgcctc tgctgcaccg tttcctttcc	2400
	tcccagctcc ttccttctac tctccatgat aacctttaag caagctattg aatttttgga	2460
	aacagaaaatt aaaaaaaaaa tctgaaaagt tcttaagttt aatctttggt taataatgaa	2520
	gtggagaacg catacaagtt tatgtatttt ttctcatctc cacataattg tattttttct	2580
	ctaagtatgt ttcaaatgat acaaaatata tactttatca attatctgat caaattgatg	2640
30	aatttttgag ctttgacgtg ttaggtctat ctaataaacg tagtaacgaa tttgggtttg	2700
	gaaatgaaat ccgataaccg atgatgggtg agagttaaac gattaaaccg ggttggttaa	2760
	aggtctcgag tctcgacggc tgcggaatc ggaaaatcac gattgaggac tttgagctgc	2820
	cacgaagatg gcgatgaggt tgaaatcaat	2850
35	<210> 94	
	<211> 3660	
	<212> DNA	
	<213> Arabidopsis sp	

&lt;400&gt; 94

	tatttgtatt tttattgtta aattttatga tttcaccggg tatatatcat cccatattaa	60
	tatttagattt attttttggg ctttatttgg gttttcgatt taaactgggc ccattctgct	120
	tcaatgaaac cctaattgggt tttgtttggg ctttggattt aaaccgggccc cattctgctt	180
5	caatgaagggt cttttgtcca acaaaaactaa catccgacac aactagtatt gccaaagagga	240
	tcgtgccaca tggcagttat tgaatcaaag gccgccaaaa ctgtaacgta gacattactt	300
	atctccggta acggacaacc actcgtttcc cgaaacagca actcacagac tcacaccact	360
	ccagtctccg gcttaactac caccagagac gattctctct tccgtcgggt ctatgacttc	420
	gattctcaac actgtctcca ccatccactc ttccagagtt acctccgtcg atcgagtcgg	480
10	agtcctctct cttcggaatt cggattccgt tgagttcact cgccggcggt ctggtttctc	540
	gacgttgatc tacgaatcac ccggtagtta gcattctggt ggatagattg atgaatgttt	600
	tcttcgattt tttttttact gatcttggtg tggatctctc gtagggcgga gatttggtgt	660
	gcgtgcggcg gagactgata ctgataaagg tatgattttt tagttgtttt tattttctct	720
	ctcttcaaaa ttctcttttc aaacactgtg gcgtttgaat ttccgacggc agttaaatct	780
15	cagacacctg acaaggcacc agccgggtgt tcaagcatta accagcttct cggtatcaaa	840
	ggagcatctc aagaaaactgt aattttgttc atctcctcag aatcttttaa attatcatat	900
	ttgtggataa tgatgtgtta gtttaggaat ttctacta aaggtaatct cttttgagga	960
	caagtcttgt ttttagctta gaaatgatgt gaaaatgttg tttgttagct aaaaagagtt	1020
	tgttgttata ttctgtattc agaataaatg gaagattcgt cttcagctta caaaaccagt	1080
20	cacttggcct ccactgggtt ggggagtcgt ctgtgggtgct gctgcttcag gtaatcatat	1140
	gaacctcttt tggatcatgc aatactgtac agaaagtttt ttcatcttcc ttccaattgt	1200
	ttcttctggc agggaaacttt cattggaccc cagaggatgt tgctaagtcg attctttgca	1260
	tgatgatgtc tggtccttgt cttactggct atacacaggt ctggttttac acaacaaaaa	1320
	gctgacttgt tcttattcta gtgcatttgc ttgggtgctac aataacctag acttgctgat	1380
25	ttccagacaa tcaacgactg gtatgataga gatatcgacg caattaatga gccatatcgt	1440
	ccaattccat ctggagcaat atcagagcca gaggttaactg agacagaaca ttgtgagctt	1500
	ttatctcttt tgtgattctg atttctcctt actccttaaa atgcagggtta ttacacaagt	1560
	ctgggtgcta ttattgggag gtcttggtat tgctggaata ttagatgtgt gggtaagtgt	1620
	gcccttctga cattaactag tacagttaaa gggcacatca gatttgctaa aatcttccct	1680
30	tatcaggcag ggcataccac tcccactgtc ttctatcttg ctttgggagg atcattgcta	1740
	tcttatatat actctgctcc acctcttaag gtaagtttta ttcttaactt ccactctcta	1800
	gtgataagac actccatcca agttttggag ttttgaatat cgatatctga actgatctca	1860
	ttgcagctaa aacaaaatgg atgggttggg aattttgcac ttggagcaag ctatattagt	1920
	ttgccatggt aagatatctc gtgtatcaat aatatatggc gttgttctca tctcattgat	1980
35	ttgtttcttg ctacttgac tgataggtgg gctggccaag cattgtttgg cactcttacg	2040
	ccagatgttg ttgttctaac actcttgtag agcatagctg gggtagctct ttggcaaacc	2100
	ttttatgttg cttttttcgt tatctgttgt aatatgctct tgcttcatgt tgtaccttg	2160
	tgataatgca gtttaggaata gccattgtta acgacttcaa aagtgttgaa ggagatagag	2220
	cattaggact tcagtctctc ccagtagctt ttggcaccga aactgcaaaa tggatatgag	2280

	ttggtgctat agacattact cagctttctg ttgccggtat gtactatcca ctgtttttgt	2340
	gcagctgtgg cttctatttc ttttccttga tcttatcaac tggatattca ccaatggtaa	2400
	agcacaaatt aatgaagctg aatcaacaaa ggcaaaacat aaaagtacat tctaataaaa	2460
	tgagctaatt aagaggaggc atctactttt atgtttcatt agtgtgattg atggattttc	2520
5	atctcatgct tctaaaacaa gtattttcaa cagtgtcatg aaataacaga acttatatct	2580
	tcatttgtac ttttactagt ggatgagtta cacaatcatt gttatagaac caaatcaaag	2640
	gtagagatca tcattagtat atgtctatct ttggttgcagg atatctatta gcactctgga	2700
	aaccttatta tgcgttggcg ttggttgcct tgatcattcc tcagattgtg ttccaggtaa	2760
	agacgttaac agtctcacat tataattaat caaattcttg tcaactcgtc gattgctaca	2820
10	ctcgtttcta taaactgcag tttaaatact ttctcaagga cctgtcaaaa tacgacgtca	2880
	agtaccaggt aagtcaactt agtacacatg tttgtgttct tttgaaatat ctttgagagg	2940
	tctcttaate agaagttgct tgaaacactc atcttgatta caggcaagcg cgcagccatt	3000
	cttggtgctc ggaatatttg taacggcatt agcatcgcaa cactgaaaaa ggcgtatttt	3060
	gatgggggtt tgcgaaagc agaggtgttg acacatcaaa tgtgggcaag tgatggcacc	3120
15	aactagttta aaagattttg taaaatgtat gtaccgttat tactagaaac aactcctgtt	3180
	gtatcaattt agcaaaacgg ctgagaaatt gtaattgatg ttaccgtatt tgcgctccat	3240
	ttttgcattt cctgctcata tcgaggattg gggtttatgt tagttctgtc acttctctgc	3300
	tttcagaatg tttttgtttt ctgtagtga ttttaactat tttcatcact ttttgtattg	3360
	attctaaaca tgtatccaca taaaaacagt aatatacaaa aatgatactt cctcaaactt	3420
20	tttataatct aaatctaaca actagctagt aacccaacta acttcataca attaatgtga	3480
	gaaactacaa agactagact atacatatgt tatttaacaa cttgaaactg tgttattact	3540
	acctgatttt tttctattct acagccattt gatatgctgc aatcttaaca tatcaagtct	3600
	cacgttggtg gacacaacat actatcacia gtaagacacg aagtaaaacc aaccggcaac	3660
25	<210> 95	
	<211> 1236	
	<212> DNA	
	<213> SOY	
30	<400> 95	
	atggattcac tgcttcttcg atctttccct aatattaata acgcctcttc tctcaccacc	60
	actggtgcaa atttctccag gactaaatct ttgcgaaca tttaccatgc aagttcttat	120
35	gtgcaaatg cttcatggca caataggaaa atccaaaaag aatataattt tttgaggttt	180
	cgggtggcaa gtttgaacca tcattacaaa ggcattgagg gagcgtgtac atgtaaaaaa	240
	tgtaataata aatttgttgt gaaagcgacc tctgaaaaat ctcttgagtc tgaacctcaa	300
40	gcttttgatc caaaaagcat tttggactct gtcaagaatt ccttggatgc tttctacagg	360
	ttttccaggc ctacacagct tattggcaca gcattaagca taatttctgt gtctcttctt	420

	gctgttgaga aaatatcaga tatatctcca ttatTTTTTA ctgggtgtgtt ggaggctgtg	480
5	gttgctgccc tgtttatgaa tatttatatt gttggtttga atcaattgtc tgatgttgaa	540
	atagacaaga taaacaagcc gtatcttcca ttagcatctg gggaatattc ctttgaaact	600
	ggtgtcacta ttgttgcac tttttcaatt ctgagttttt ggcttggctg ggtttaggt	660
10	tcattggccat tattttgggc cttttttgta agctttgtgc taggaactgc ttattcaatc	720
	aatgtgcctc tgttgagatg gaagaggttt gcagtgttg cagcgatgtg cattctagct	780
15	gttcgggcag taatagttca acttgcattt ttccttcaca tgcagactca tgtgtacaag	840
	aggecacctg tcttttcaag accattgatt tttgtactg cattcatgag cttcttctct	900
	gtagttatag cactgtttta ggatatacct gacattgaag gagataaagt atttggcatc	960
20	caatcttttt cagtgcgttt aggtcagaag ccggtgttct ggacttgtgt tacccttctt	1020
	gaaatagctt atggagtgc cctcctgggt ggagctgcat ctcttgtct ttggagcaaa	1080
25	attttcacgg gtctgggaca cgtgtgtgtg gcttcaattc tctggtttca tgccaaatct	1140
	gtagatttga aaagcaaagc ttcgataaca tccttctata tgtttatttg gaagctattt	1200
	tatgcagaat acttactcat tccttttgtt agatga	1236
30		
	<210> 96	
	<211> 1188	
	<212> DNA	
	<213> SOY	
35		
	<400> 96	
	atggattcga tgcttcttcg atcttttctt aatattaaca acgcttcttc tctcgccacc	60
40	actggttctt atttgccaaa tgcttcatgg cacaatagga aaatccaaaa agaataaat	120
	tttttgaggt ttcgggtggcc aagtttgaac caccattaca aaagcattga aggagggtgt	180
	acatgtaaaa aatgtaatat aaaatttgtt gtgaaagcga cctctgaaaa atcttttgag	240
45	tctgaacccc aagcttttga tccaaaaagc attttggact ctgtcaagaa ttccttgat	300
	gctttctaca ggttttccag acctcacaca gttattggca cagcattaag cataatttct	360
50	gtgtccctcc ttgtgttgga gaaaatatca gatatatctc cattatTTTT tactgggtgtg	420
	ttggaggctg tggttgctgc cctgtttatg aatatttata ttgttggttt gaatcaattg	480
	tctgatgttg aaatagacaa gataaacaag ccgtatcttc cattagcatc tggggaatat	540

tcctttgaaa ctggtgtcac tattgttgca tctttttcaa ttctgagttt ttggcttggc 600  
 tgggtttag gttcatggcc attattttgg gccctttttg taagctttgt gctaggaact 660  
 5 gcttattcaa tcaatgtgcc tctgttgaga tggaagaggt ttgcagtgc tgcagcgatg 720  
 tgcattctag ctgttcgggc agtaatagtt caacttgcac ttttccttca catccagact 780  
 10 catgtataca agaggccacc tgtcttttca agatcattga tttttgctac tgcattcatg 840  
 agcttcttct ctgtagttat agcactgttt aaggatatac ctgacattga aggagataaa 900  
 gtatttggca tccaatcttt ttcagtgcgt ttaggtcaga agccggtatt ctggacttgt 960  
 15 gttatccttc ttgaaatagc ttatggagtc gccctcctgg tgggagctgc atctccttgt 1020  
 ctttggagca aaattgtcac gggctctggga cacgctgttc tggcttcaat tctctggttt 1080  
 20 catgccaaat ctgtagattt gaaaagcaaa gcttcgataa catccttcta tatgtttatt 1140  
 tggaagctat tttatgcaga atacttactc attccttttg ttagatga 1188

<210> 97  
 25 <211> 395  
 <212> PRT  
 <213> SOY

<400> 97  
 30

Met Asp Ser Met Leu Leu Arg Ser Phe Pro Asn Ile Asn Asn Ala Ser  
 1 5 10 15  
 35 Ser Leu Ala Thr Thr Gly Ser Tyr Leu Pro Asn Ala Ser Trp His Asn  
 20 25 30  
 Arg Lys Ile Gln Lys Glu Tyr Asn Phe Leu Arg Phe Arg Trp Pro Ser  
 35 40 45  
 40 Leu Asn His His Tyr Lys Ser Ile Glu Gly Gly Cys Thr Cys Lys Lys  
 50 55 60  
 45 Cys Asn Ile Lys Phe Val Val Lys Ala Thr Ser Glu Lys Ser Phe Glu  
 65 70 75 80  
 Ser Glu Pro Gln Ala Phe Asp Pro Lys Ser Ile Leu Asp Ser Val Lys  
 85 90 95  
 50 Asn Ser Leu Asp Ala Phe Tyr Arg Phe Ser Arg Pro His Thr Val Ile  
 100 105 110  
 Gly Thr Ala Leu Ser Ile Ile Ser Val Ser Leu Leu Ala Val Glu Lys  
 115 120 125

Ile Ser Asp Ile Ser Pro Leu Phe Phe Thr Gly Val Leu Glu Ala Val  
 130 135 140  
 5 Val Ala Ala Leu Phe Met Asn Ile Tyr Ile Val Gly Leu Asn Gln Leu  
 145 150 155 160  
 Ser Asp Val Glu Ile Asp Lys Ile Asn Lys Pro Tyr Leu Pro Leu Ala  
 165 170 175  
 10 Ser Gly Glu Tyr Ser Phe Glu Thr Gly Val Thr Ile Val Ala Ser Phe  
 180 185 190  
 Ser Ile Leu Ser Phe Trp Leu Gly Trp Val Val Gly Ser Trp Pro Leu  
 195 200 205  
 Phe Trp Ala Leu Phe Val Ser Phe Val Leu Gly Thr Ala Tyr Ser Ile  
 210 215 220  
 20 Asn Val Pro Leu Leu Arg Trp Lys Arg Phe Ala Val Leu Ala Ala Met  
 225 230 235 240  
 Cys Ile Leu Ala Val Arg Ala Val Ile Val Gln Leu Ala Phe Phe Leu  
 245 250 255  
 25 His Ile Gln Thr His Val Tyr Lys Arg Pro Pro Val Phe Ser Arg Ser  
 260 265 270  
 Leu Ile Phe Ala Thr Ala Phe Met Ser Phe Phe Ser Val Val Ile Ala  
 275 280 285  
 30 Leu Phe Lys Asp Ile Pro Asp Ile Glu Gly Asp Lys Val Phe Gly Ile  
 290 295 300  
 35 Gln Ser Phe Ser Val Arg Leu Gly Gln Lys Pro Val Phe Trp Thr Cys  
 305 310 315 320  
 Val Ile Leu Leu Glu Ile Ala Tyr Gly Val Ala Leu Leu Val Gly Ala  
 325 330 335  
 40 Ala Ser Pro Cys Leu Trp Ser Lys Ile Val Thr Gly Leu Gly His Ala  
 340 345 350  
 Val Leu Ala Ser Ile Leu Trp Phe His Ala Lys Ser Val Asp Leu Lys  
 355 360 365  
 45 Ser Lys Ala Ser Ile Thr Ser Phe Tyr Met Phe Ile Trp Lys Leu Phe  
 370 375 380  
 50 Tyr Ala Glu Tyr Leu Leu Ile Pro Phe Val Arg  
 385 390 395  
 <210> 98  
 <211> 411  
 55 <212> PRT  
 <213> SOY

&lt;400&gt; 98

5 Met Asp Ser Leu Leu Leu Arg Ser Phe Pro Asn Ile Asn Asn Ala Ser  
 1 5 10 15  
 Ser Leu Thr Thr Thr Gly Ala Asn Phe Ser Arg Thr Lys Ser Phe Ala  
 20 25 30  
 10 Asn Ile Tyr His Ala Ser Ser Tyr Val Pro Asn Ala Ser Trp His Asn  
 35 40 45  
 Arg Lys Ile Gln Lys Glu Tyr Asn Phe Leu Arg Phe Arg Trp Pro Ser  
 50 55 60  
 15 Leu Asn His His Tyr Lys Gly Ile Glu Gly Ala Cys Thr Cys Lys Lys  
 65 70 75 80  
 Cys Asn Ile Lys Phe Val Val Lys Ala Thr Ser Glu Lys Ser Leu Glu  
 85 90 95  
 Ser Glu Pro Gln Ala Phe Asp Pro Lys Ser Ile Leu Asp Ser Val Lys  
 100 105 110  
 25 Asn Ser Leu Asp Ala Phe Tyr Arg Phe Ser Arg Pro His Thr Val Ile  
 115 120 125  
 Gly Thr Ala Leu Ser Ile Ile Ser Val Ser Leu Leu Ala Val Glu Lys  
 130 135 140  
 30 Ile Ser Asp Ile Ser Pro Leu Phe Phe Thr Gly Val Leu Glu Ala Val  
 145 150 155 160  
 Val Ala Ala Leu Phe Met Asn Ile Tyr Ile Val Gly Leu Asn Gln Leu  
 165 170 175  
 Ser Asp Val Glu Ile Asp Lys Ile Asn Lys Pro Tyr Leu Pro Leu Ala  
 180 185 190  
 40 Ser Gly Glu Tyr Ser Phe Glu Thr Gly Val Thr Ile Val Ala Ser Phe  
 195 200 205  
 Ser Ile Leu Ser Phe Trp Leu Gly Trp Val Val Gly Ser Trp Pro Leu  
 210 215 220  
 45 Phe Trp Ala Leu Phe Val Ser Phe Val Leu Gly Thr Ala Tyr Ser Ile  
 225 230 235 240  
 Asn Val Pro Leu Leu Arg Trp Lys Arg Phe Ala Val Leu Ala Ala Met  
 245 250 255  
 Cys Ile Leu Ala Val Arg Ala Val Ile Val Gln Leu Ala Phe Phe Leu  
 260 265 270  
 55 His Met Gln Thr His Val Tyr Lys Arg Pro Pro Val Phe Ser Arg Pro  
 275 280 285

Leu Ile Phe Ala Thr Ala Phe Met Ser Phe Phe Ser Val Val Ile Ala  
 290 295 300  
 5 Leu Phe Lys Asp Ile Pro Asp Ile Glu Gly Asp Lys Val Phe Gly Ile  
 305 310 315 320  
 Gln Ser Phe Ser Val Arg Leu Gly Gln Lys Pro Val Phe Trp Thr Cys  
 325 330 335  
 10 Val Thr Leu Leu Glu Ile Ala Tyr Gly Val Ala Leu Leu Val Gly Ala  
 340 345 350  
 Ala Ser Pro Cys Leu Trp Ser Lys Ile Phe Thr Gly Leu Gly His Ala  
 15 355 360 365  
 Val Leu Ala Ser Ile Leu Trp Phe His Ala Lys Ser Val Asp Leu Lys  
 370 375 380  
 20 Ser Lys Ala Ser Ile Thr Ser Phe Tyr Met Phe Ile Trp Lys Leu Phe  
 385 390 395 400  
 Tyr Ala Glu Tyr Leu Leu Ile Pro Phe Val Arg  
 405 410  
 25  
 <210> 99  
 <211> 964  
 <212> DNA  
 <213> RICE  
 30  
 <400> 99  
 gagcagcact gggctcttaca ttccaatgga gctcgccctgt tgcttttcatt acatgcttctg 60  
 tgacttttatt tgcttttggtc attgctataa ccaaagatct cccagatggt gaaggggatc 120  
 35 ggaagtatca aatatcaact ttggcgacaa agctcgggtgt cagaaacatt gcatttcttg 180  
 gctctgggtt attgatagca aattatgttg ctgctattgc ttagctttt ctcatgcctc 240  
 40 aggctttcag gcgcactgta atgggtgcctg tgcattgctgc ccttgccgtt ggtataattt 300  
 tccagacatg ggttctggag caagcaaaat atactaagga tgctatttca cagtactacc 360  
 ggttcatttg gaatctcttc tatgctgaat acatcttctt cccgttgata tagagaccaa 420  
 45 gcaatctgat atgggtctgca tggtgagtgc ggcaaaaact agaagcccat atgaacagtg 480  
 ggagtagggg aacgaacatg ccatccatgg gaagactctg ataactctct ctgccccggg 540  
 50 ctgtaaaggg taagcactgt tgggcatata tatgaaagga aggtgataaa gcagggatgc 600  
 taaattgcta ctgggatcct caaaggctta tagtggtcac cagtggaatg tgccttaata 660  
 atttggttac ccagcagagc aagtttttgc aggttattag gtaatatctt tgagggaatg 720



aacttagatt tcattgtttt aaggtctggt cacacaacgg gtagtagtgc tggagcggca 780  
 5 aaaaacgacc ttgttttaca ctaccaaggg aggttaactc tagttttcat gtgaccactt 840  
 accttgagag ttgagaccat ggaatcactt gtcgactcct cggcttgtat atttctagtg 900  
 tcagcatttg cattctcctc cccacttgta cttgaaaagt tgaagacaac ttttttgttt 960  
 10 gtgt 964

<210> 100  
 <211> 421  
 15 <212> DNA  
 <213> WHEAT

<400> 100  
 20 cgtccgcgga cgcgtgggtg cttattcagt caatctgccg cactttctat ggaagagatc 60  
 tgctgttggt gcagcactct gcatattagc agtgcgtagc gtgatagtcc aactggcatt 120  
 ttttctccac attcagacat ttgttttcag aaggccggca gacttttcaa agccattgat 180  
 25 atttgcaact gccttcatga cattcttctc agttgtaata gcattattca aggatatacc 240  
 cgatattgaa ggggaccgca tctttggaat ccaatctttt agtggtagac taggtcaaag 300  
 30 caggggtttc tggacttgcg ttggcctact tgaggttgcc tacggtgttg cgatactgag 360  
 gggggtaact tcttccagtt tgtggagcaa atctataact gttgtgggcc atgcaatcct 420  
 c 421

35  
 <210> 101  
 <211> 705  
 <212> DNA  
 <213> LEEK

40  
 <400> 101  
 gtttcccccc ctccaatttt tttttttttt ttttacttca tttttctgtg aataaattct 60  
 45 taaaaaagac aaagaaaacc actggatate cttaaattcaa cataggctat tgtcattcaa 120  
 tgataatctt taacacaaca tacaacatga atataattaa ggagaaatga tctgcaattg 180  
 ttgaaagaac tctccgtttt taagatgaca attaaagcgt tgtaattcc agccatttct 240  
 50 gcctccatta tctactcctc ttctcttgcg attcttttcc atgtaggtca taaaccctca 300

tcttacaaaa ggaatgagca agtactcagc atagaagagc ttccacacga acatataaaa 360  
 5 agatgtaata gtggttttgg tcattggtcc atgagatcta gcacgattcc aaagtaacga 420  
 cccaagaatt gcatgaccta tcaactgttaa gcatttgctc cataggcatg aggaagtagc 480  
 tccaacaacc atgacaacag tgtaggcat ctcaaggaga tatatacata tccaaaacac 540  
 10 cctctcctgg ccaaggcgca cgctgaaaga atggatgcca aatattttgt ctccgtctat 600  
 atcaggata tccttaaata gagcaataac aactgagaag aagctcatga aggcagttgc 660  
 15 aaatatcaat ggccttgtga aacttgctgg tcttttgaaa acaaa 705

<210> 102  
 <211> 637  
 <212> DNA  
 20 <213> LEEK

<220>  
 <221> misc\_feature  
 <222> (1)..(637)  
 25 <223> n = g, a, t or c

<400> 102  
 30 nattcggcac gagttttgaa gaagttaagc atggactccc tccttaccaa gccagttgta 60  
 atacctctgc cttctccagt ttgttcaact ccaatcttgc gaggcagttc tgcaccaggg 120  
 cagtattcat gtagaaacta caatccaata agaattcaaa ggtgcctcgt aaattatgaa 180  
 35 catgtgaaac caaggtttac aacatgtagt aggtctcaaa aacttggtca tgtaaaagcc 240  
 acatccgagc attctttaga atctggatcc gaaggataca ctctagaag catatgggaa 300  
 gccgtactag cttcactgaa tgttctatac aaattttcac gacctcacac aataatagga 360  
 40 acagcaatgg gcataatgtc agtttctttg cttgttgctg agagcctatc cgatatttct 420  
 cctctgtttt ttgtgggatt attagaggct gtggttgctg cattgtttat gaatgtttac 480  
 45 attgtaggtc tgaatcaatt atttgacata gaaatagaca aggtcaataa acctgatctt 540  
 cctcttgcat ctggagaata ctaccaaga gctggtagct ctattgtcat tgcttcagcc 600  
 50 atcatgagct ttggcattgg atggttagtt ggctctt 637

<210> 103  
 <211> 677  
 <212> DNA  
 5 <213> CANOLA

<400> 103  
 tttttttttt tttttttcaa aaagaccaat ccttttagtat gtacatgaac aaagtgattt 60  
 10 tgtctccaag ctacaaagaa gaagaagaga ggtatacaaa gaaaactaca aatgttcacc 120  
 atgaatgcta gaagaagggg aataacagat actctgcgta gaagagattc catataaacc 180  
 15 ggtaatatcc tgctatagct tcctttgtgt agtttgcttt ttctagcacc catgtctgga 240  
 aaaccaagca tgaagccaag atcatatgtg caggaatcat caagctacct ctaaaaacct 300  
 gaggcattga gaaagctagt gatatggcag aaatatagtt cactagcaga agtccagaac 360  
 20 cgaggaatgc aatgttcctc actccaagct ttgttgctag tggtgatatt tggaacttgc 420  
 gatctccttc aacatcagga agatcttttg taatagcaat gactagtga aacagtgtca 480  
 25 caaaagacgt gatgaaagcc acaggtgcac tccactgaaa cgaaagtcca agagcagctc 540  
 tagtagcatg gtacacacca aaattaagaa gaaaacctcg taccgtggca ataataagaa 600  
 acgctgcaac tggaaatctc ttcattctaa atggtggaac agaatagatg gtccccagat 660  
 30 cggacgcgtg ggtcgac 677

<210> 104  
 <211> 1431  
 35 <212> DNA  
 <213> CORN

<400> 104  
 40 ccacgcgtcc gcccggccaa gggatggacg cgcttcgcct acggccgtcc ctctccccg 60  
 tgcggccccg gcgggccgc ccgcgagatc attttctacc accatgttgt tccatacaac 120  
 gaaatggtga aggacgaatt tgcttttcta gccaaaggac ccaaggteet accttgcatc 180  
 45 accatcagaa attcttcgaa tggaaatcct cctattgtag gatatcacat cggtcattaa 240  
 atacttctgt taatgcttcg gggcaacagc tgcagtctga acctgaaaca catgattcta 300  
 50 caaccatctg gagggcaata tcattctctc tagatgcatt ttacagattt tccccgccac 360

atactgtcat aggaacagca ttaagcatag tctcagtttc ccttctagct gtccagagct 420  
 tgtctgatat atcacctttg ttcctcactg gtttgctgga ggcagtggta gctgcccttt 480  
 5 tcatgaatat ctatatgtt ggactgaacc agttattcga cattgagata gacaaggtta 540  
 acaagccaac tcttccattg gcatctgggg aatacaccct tgcaactggg gttgcaatag 600  
 10 tttcgggtctt tgccgctatg agctttggcc ttggatgggc tgttggatca caacctctgt 660  
 tttgggctct tttcataagc tttgttcttg ggactgcata ttcaatcaat ctgccgtacc 720  
 ttcgatggaa gagatttgct gttgttgag cactgtgcat attagcagtt cgtgcagtga 780  
 15 ttgttcagct ggcctttttt ctccacattc agacttttgt tttcaggaga ccggcagtgt 840  
 tttctaggcc attattattt gcaactggat ttatgacgtt cttctctgtt gtaatagcac 900  
 tattcaagga tatacctgac atcgaagggg accgcatatt cgggatccga tccctcagcg 960  
 20 tccggttagg gcaaaagaag gtcttttgga tctgcgttgg cttgcttgag atggcctaca 1020  
 gcgttgcat actgatggga gctacctctt cctgtttgtg gagcaaaaca gcaaccatcg 1080  
 25 ctggccattc catacttgcc gcgacccat ggagctgcgc gcgacgggtg gacttgacga 1140  
 gcaaagccgc aataacgtcc ttctacatgt tcatctggaa gctgttctac gcggagtacc 1200  
 30 tgctcatccc tctgggtgcgg tgagcgcgag gcgaggtggt ggcagacgga tcggcgtcgg 1260  
 cggggcggca aacaactcca cgggagaact tgagtgcgg aagtaaaactc ccgtttgaaa 1320  
 gttgaagcgt gcaccaccgg caccgggcag agagagacac ggtggctgga tggatacggg 1380  
 35 tggccccccc aataaattcc cccgtgcatg gtaaaaaaaaa aaaaaaaaaa a 1431

&lt;210&gt; 105

&lt;211&gt; 1870

40 &lt;212&gt; DNA

&lt;213&gt; CORN

&lt;400&gt; 105

45 gccgcgcagc gcgacgagcg ccacctgctt gctgccgcgt gcctgcgtgc gtgtgcgtcc 60  
 accactgacc ccgcgcccgc ccgcgcccc tgcccccca ctccacttgc tcaactcgtc 120  
 cggcccgtt cccccccggc caagggatgg acgcgcttcg cctacggcgg tccctcctcc 180  
 50 ccgtgcggcc cggcgcggcc cggcgcgag atcattttct accaccatgt tgttccatac 240  
 aacgaaatgg tgaaggacga atttgctttt ctagccaaag gacccaaggt cctaccttgc 300  
 atcaccatca gaaattcttc gaatggaaat cctcctattg taggatatca catcggtcat 360

	taaataacttc tgttaatgct tcggggcaac agctgcagtc tgaacctgaa acacatgatt	420
5	ctacaaccat ctggagggca atatcatctt ctctagatgc attttacaga ttttcccggc	480
	cacatactgt cataggaaca gcattaagca tagtctcagt ttcccttcta gctgtccaga	540
	gcttgtctga tatatcacct ttgttcctca ctggtttgct ggaggcagtg gtagctgccc	600
10	ttttcatgaa tatctatatt gttggactga accagttatt cgacattgag atagacaagg	660
	ttaacaagcc aactcttcca ttggcatctg gggaatacac ccttgcaact ggggttgcaa	720
15	tagtttcggc ctttgccgct atgagctttg gccttgatg ggctgttgga tcacaacctc	780
	tgttttgggc tcttttcata agctttgttc ttgggactgc atattcaatc aatctgccgt	840
	accttcgatg gaagagattt gctgttggtg cagcactgtg catattagca gttcgtgcag	900
20	tgattgttca gctggccttt tttctccaca ttcagacttt tgttttcagg agaccggcag	960
	tgttttctag gccattatta ttgcaactg gatttatgac gttcttctct gttgtaatag	1020
25	cactattcaa ggatatacct gacatcgaag gggaccgcat attcgggac cgccttca	1080
	gcgtccgggt agggcaaaag aaggctcttt ggatctgcgt tggcttgctt gagatggcct	1140
	acagcgttgc gatactgatg ggagctacct ctccctgttt gtggagcaaa acagcaacca	1200
30	tcgctggcca ttccatactt gccgcgatcc tatggagctg cgcgcgatcg gtggacttga	1260
	cgagcaaagc cgcaataacg tccttctaca tgttcatctg gaagctgttc tacgcggagt	1320
35	acctgctcat ccctctgggtg cggtgagcgc gaggcgaggt ggtggcagac ggatcggcgt	1380
	cggcggggcg gcaaacaact ccacgggaga acttgagtgc cggaagtaaa ctcccgtttg	1440
	aaagttgaag cgtgcaccac cggcaccggg cagagagaga cacggtggct ggatggatac	1500
40	ggatggcccc cccaataaat tccccgtgc atggtacccc acgctgcttg atgatatccc	1560
	atgtgtccgg gtgaccggac ctgatcgtct ctagagagat tggttgcaca acgtccaaca	1620
45	tagcccgtag gtattgctac cactgctagt atgatactcc ttccctagtcc ttgccagcac	1680
	cagtgaacca aacttggtcg gctgagctca gcgctcagca gctttacgtg catctgcgcc	1740
	ttgacttgtg cagtgggcgt cgctagcatg aatgatgtat ggtgcgtcac ggcctgacgg	1800
50	ttcgtcagtc tgggcccgtgt tttgtgtccg aggaagatcg tctgtcagag atctggattg	1860
	cctcgtgct	1870
55	<210> 106	
	<211> 642	

WO 01/79472

PCT/US01/12334

<212> DNA

<213> CORN

<400> 106

5 cggccggact cttctgactt ggcaaccgcc ggcagcgcg acgagcgcca cctgcttget 60  
 gccgcgtgcc tgcgtgcgtg tgcgtccacc actgaccccg cgcccgcgccc cggcccctgc 120  
 10 ccctccactc cacttgctca ctgcgtcggct cgtcgcggcc cgcttcccc cgggccaagg 180  
 gatggacgcg cttcgccctac ggccgtccct cctccccgtg cggcccggcg cggcccgcgc 240  
 gcgaggcagt ggtagctgcc cttttcatga atatctatat tgttggactg aaccagttat 300  
 15 tcgacattga gatagacaag gttaacaagc caactcttcc attggcatct ggggaatata 360  
 cccttgcaac tgggggttgca atagtttcgg tctttgcgc tatgagcttt ggccttggat 420  
 gggctgttgg atcacaacct ctgttttggg ctcttttcat aagctttgtt cttgggactg 480  
 20 catattcaat caatctgccg taccttcgat ggaagagatt tgcgtgtgtt gcagcactgt 540  
 gcatattage agttcgtgca gtgattgttc agctggcctt tttctccac attcagactt 600  
 25 ttgttttcag gagaccggca gtgttttcta ggccattatt at 642

<210> 107

<211> 362

30 <212> DNA

<213> COTTON

<400> 107

35 cccacgcgtc cgaacattgt ttgcacttgt tattgccata accaaggatc ttccagatgt 60  
 agaaggagat cgcaaatttc aaatatcaac attagcaaca aagcttggag ttagaaatat 120  
 tgcatttctt ggttccggac ttctactggt gaattatgtt gctgctgtgt tggetgcaat 180  
 40 atacatgcct caggctttca ggcgtagttt aatgatacct gtcataatct ttttggcggt 240  
 ctgcttgatt tttcagacat ggggtgttga acaagcaaata taaaaaagg aagcaatctc 300  
 ggggttctat cgtttcatat ggaatctctt ctatgcagag tatgcgattt tccccctcgt 360  
 45 gt 362

<210> 108

50 <211> 575

&lt;212&gt; DNA

&lt;213&gt; TOMATO

&lt;400&gt; 108

5 cagatcaatt ccagttcctg ctgagttttc tccactcaaa accagttcac atgcaatagt 60  
 acggggttttg aaatgtaaag catggaagag accaaaaaag cactattcct cttcaatgaa 120  
 gttgcagcgg cagtatatca cgcaagagca tggtggagga agtgatctaa gcactattgc 180  
 10 tgctgataaa aaacttaaag ggagatTTTT ggtgcacgca tcatctgaac accctcttga 240  
 atctcaacct tctaaaagtc cttgggactc agttaatgat gccgtagatg ctttctacag 300  
 15 gttctcgcg ccccatacca taataggaac agcattgagc ataatttcag tttctctcct 360  
 tgcagttgag aagttctctg atttttctcc attatttttc actgggggtg tagaggccat 420  
 20 tggtgctgcc ctattcatga acatttacat agttggttta aaccagttgt ctgacatcga 480  
 aatagacaag gtaaacaagc catatcttcc attggcatca ggggaatact ctgtacaaac 540  
 tggagtgatt gttgtgtcgt cttttgccat tttga 575

25

&lt;210&gt; 109

&lt;211&gt; 1663

&lt;212&gt; DNA

&lt;213&gt; ARABIDOPSIS

30

&lt;400&gt; 109

aacaccaaac acacaatttc acattctttt gcatatttct tcttcttctt ccattatgga 60  
 gatacggagc ttgattgttt ctatgaaccc taatttatct tcctttgagc tctctcgccc 120  
 35 tgtatctcct ctactcgct cactagtccc gttccgatcg actaaactag ttccccgctc 180  
 catttctagg gggatcccgt cgatctccac cccgaatagt gaaactgaca agatctccgt 240  
 40 taaacctgtt tacgtcccga cgtctcccaa tcgcgaactc cggactcctc acagtggata 300  
 ccatttcgat ggaacacctc ggaagttctt cgagggatgg tggatccggg tttccatccc 360  
 agagaagagg gagagttttt gttttatgta ttctgtggag aatcctgcat ttcggcagag 420  
 45 tttgtcacca ttggaagtgg ctctatatgg acctagattc actggtgttg gagctcagat 480  
 tcttggcgct aatgataaat atttatgcca atacgaacaa gactctcaca atttctgggg 540  
 50 agatcgacat gagctagttt tggggaatac ttttagtgct gtgccaggcg caaaggctcc 600  
 aaacaaggag gttccaccag aggaatttaa cagaagagtg tccgaagggt tccaagctac 660

tccatTTTtgg catcaaggTc acattTtgcga tgatggccgt actgactatg cggaaactgt 720  
 5 gaaatctgct cgttgggagt atagtactcg tcccgtttac ggTtggggTg atgttggggc 780  
 caaacagaag tcaactgcag gctggcctgc agcttttccct gtatttgagc ctcattggca 840  
 gatatgcatg gcaggaggcc tttccacagg gtggatagaa tggggcggtg aaaggTttga 900  
 10 gtttcgggat gcaccttctt attcagagaa gaattggggT ggaggcttcc caagaaaatg 960  
 gttttgggtc cagtgtaatg tctttgaagg ggcaactgga gaagtTgctt taaccgcagg 1020  
 15 tggcgggTtg aggcaattgc ctggattgac tgagacctat gaaaatgctg cactggTttg 1080  
 tgtacactat gatggaaaaa tgtacgagtt tgttccttgg aatggTgTtg ttagatggga 1140  
 aatgtctccc tggggTtatt ggtatataac tgcagagAAC gaaaaccatg tggTggaact 1200  
 20 agaggcaaga acaaatgaag cgggtacacc tctgcgtgct cctaccacag aagtTgggct 1260  
 agctacggct tgcagagata gttgttacgg tgaattgaag ttgcagatat gggaacggct 1320  
 atatgatgga agtaaaggca aggtgatatt agagacaaag agctcaatgg cagcagtgga 1380  
 25 gataggagga ggaccgtggt ttgggacatg gaaaggagat acgagcaaca cgcccgagct 1440  
 actaaaacag gctcttcagg tcccatTtga tcttgaaagc gccttaggtt tggTcccttt 1500  
 30 cttcaagcca ccgggtctgt aacattgatg agtgttttgt ttgttgatag agacccatgt 1560  
 gatgaatgaa gccttagtca tgtcattgct agcttcacta ttatgtatgt atgattttag 1620  
 35 ttcgttcggT ccttgtggta aatgatacgg gccagtgtaa agt 1663

&lt;210&gt; 110

&lt;211&gt; 488

&lt;212&gt; PRT

40 &lt;213&gt; ARABIDOPSIS

&lt;400&gt; 110

45 Met Glu Ile Arg Ser Leu Ile Val Ser Met Asn Pro Asn Leu Ser Ser  
 1 5 10 15  
 Phe Glu Leu Ser Arg Pro Val Ser Pro Leu Thr Arg Ser Leu Val Pro  
 20 25 30  
 50 Phe Arg Ser Thr Lys Leu Val Pro Arg Ser Ile Ser Arg Val Ser Ala  
 35 40 45  
 Ser Ile Ser Thr Pro Asn Ser Glu Thr Asp Lys Ile Ser Val Lys Pro



	50	55	60
	Val Tyr Val Pro Thr Ser Pro Asn Arg Glu Leu Arg Thr Pro His Ser		
	65	70	75 80
5	Gly Tyr His Phe Asp Gly Thr Pro Arg Lys Phe Phe Glu Gly Trp Tyr		
		85	90 95
	Phe Arg Val Ser Ile Pro Glu Lys Arg Glu Ser Phe Cys Phe Met Tyr		
10		100	105 110
	Ser Val Glu Asn Pro Ala Phe Arg Gln Ser Leu Ser Pro Leu Glu Val		
		115	120 125
15	Ala Leu Tyr Gly Pro Arg Phe Thr Gly Val Gly Ala Gln Ile Leu Gly		
		130	135 140
	Ala Asn Asp Lys Tyr Leu Cys Gln Tyr Glu Gln Asp Ser His Asn Phe		
20		145	150 155 160
	Trp Gly Asp Arg His Glu Leu Val Leu Gly Asn Thr Phe Ser Ala Val		
		165	170 175
25	Pro Gly Ala Lys Ala Pro Asn Lys Glu Val Pro Pro Glu Glu Phe Asn		
		180	185 190
	Arg Arg Val Ser Glu Gly Phe Gln Ala Thr Pro Phe Trp His Gln Gly		
		195	200 205
30	His Ile Cys Asp Asp Gly Arg Thr Asp Tyr Ala Glu Thr Val Lys Ser		
		210	215 220
	Ala Arg Trp Glu Tyr Ser Thr Arg Pro Val Tyr Gly Trp Gly Asp Val		
35		225	230 235 240
	Gly Ala Lys Gln Lys Ser Thr Ala Gly Trp Pro Ala Ala Phe Pro Val		
		245	250 255
40	Phe Glu Pro His Trp Gln Ile Cys Met Ala Gly Gly Leu Ser Thr Gly		
		260	265 270
	Trp Ile Glu Trp Gly Gly Glu Arg Phe Glu Phe Arg Asp Ala Pro Ser		
		275	280 285
45	Tyr Ser Glu Lys Asn Trp Gly Gly Gly Phe Pro Arg Lys Trp Phe Trp		
		290	295 300
	Val Gln Cys Asn Val Phe Glu Gly Ala Thr Gly Glu Val Ala Leu Thr		
50		305	310 315 320
	Ala Gly Gly Gly Leu Arg Gln Leu Pro Gly Leu Thr Glu Thr Tyr Glu		
		325	330 335
55	Asn Ala Ala Leu Val Cys Val His Tyr Asp Gly Lys Met Tyr Glu Phe		
		340	345 350
	Val Pro Trp Asn Gly Val Val Arg Trp Glu Met Ser Pro Trp Gly Tyr		
		355	360 365

Trp Tyr Ile Thr Ala Glu Asn Glu Asn His Val Val Glu Leu Glu Ala  
370 375 380

5 Arg Thr Asn Glu Ala Gly Thr Pro Leu Arg Ala Pro Thr Thr Glu Val  
385 390 395 400

Gly Leu Ala Thr Ala Cys Arg Asp Ser Cys Tyr Gly Glu Leu Lys Leu  
405 410 415

10 Gln Ile Trp Glu Arg Leu Tyr Asp Gly Ser Lys Gly Lys Val Ile Leu  
420 425 430

15 Glu Thr Lys Ser Ser Met Ala Ala Val Glu Ile Gly Gly Gly Pro Trp  
435 440 445

Phe Gly Thr Trp Lys Gly Asp Thr Ser Asn Thr Pro Glu Leu Leu Lys  
450 455 460

20 Gln Ala Leu Gln Val Pro Leu Asp Leu Glu Ser Ala Leu Gly Leu Val  
465 470 475 480

Pro Phe Phe Lys Pro Pro Gly Leu  
485

25